INVALSI

CONFERENCE 2026
EVALUATION AND RESEARCH IN EDUCATION

In collaboration with

POLITECNICO MILANO 1863

POLO TERRITORIALE LECCO

CONFERENCE 2026

"Evaluation and Research in Education"

Demetrio Cosola, The Dictation Lesson, 1891*

Lecco, 4 - 6 November 2026

Under the patronage of

espanet ITALIA

Istat
Istituto Nazionale di Statistica

Under the support of

BE FOR EDUCATION FOUNDATION

INDEX

INTRODUCTION	**2**
ABSTRACT	**3**
AUTHORS	**335**

INTRODUCTION

We are delighted to present the Book of Abstracts for the 'Evaluation and Research in Education 2026' Conference, organised in collaboration with the Politecnico di Milano and hosted at the Lecco Campus from 4 to 6 November 2026.

The conference aims to provide a high–calibre forum for academic exchange dedicated to the analysis of education systems, fostering dialogue between the research community, policy–makers and professionals in education and training. Through the sharing of empirical evidence, innovative methodologies and research findings, the event aims to contribute to a deeper understanding of educational processes and to promote practices and policies geared towards improving the quality, effectiveness and equity of education systems.

The INVALSI databases play a particularly significant role; thanks in part to their integration with information from other agencies and institutions, they constitute an extraordinarily valuable resource for exploring key issues in educational research and for the public debate on learning opportunities and the development of human capital.

The abstracts collected in this volume bear witness to the richness and variety of theoretical, methodological and disciplinary approaches that characterise contemporary educational research. The contributions presented here offer an overview of the most relevant and topical issues in the field of education and training, fostering scientifically grounded reflection aimed at generating knowledge useful for the improvement of education systems.

We hope that reading this volume will provide an opportunity for deeper understanding and inspire new research perspectives. We wish all participants a conference rich in scholarly exchange, collaboration and professional development.

The Conference is organised by the research group of Area 2 – Statistical Service: Patrizia Falzetti (Manager), Paolo Barabanti, Andrea Bendinelli, Leonardo Boulay, Emiliano Campodifiori, Michele Cardone, Federica Collia, Ettore De Sossi, Paola Giangiacomo, Pierangelo Grosso, Jana Kopecna, Fabrizio Lasorsa, Giuseppina Le Rose, Francesca Leggi, Lorenzo Maraviglia, Michele Marsili, Giancarlo Mastrone, Alessandro Mattia, Marcello Napoli, Fiammetta Noccioli, Carlo Palmiero, Monica Papini, Daniele Rowlett, Antonio Severoni, Agnese Spoladore, Daniela Torti, Valeria F. Tortora.

Contacts: uff.statistico@INVALSI.it

The opinions and content expressed within the initiative are the sole responsibility of the speakers and are in no way attributable to INVALSI.

ABSTRACT

SESSION 2. WHAT MAKES A SCHOOL WORK? TEACHERS AND SCHOOL PRINCIPALS AS KEY ACTORS IN THE EDUCATIONAL SYSTEM

ORGANIZER: UNIVERSITY OF MILANO "BICOCCA"
COORDINATORS: GIANLUCA ARGENTIN – IVAN BLANCATO
NOVEMBER 4TH: 2.00 P.M. – 4.00 P.M. {AULA MAGNA – RESEARCH 1}

Teachers' Territorial Mobility Trajectories: Heterogeneity Across Educational and Professional Pathways

Ivan Blancato – Iacopo Moreschini

Introduction

Looking at the territorial mobility of teachers allows us to capture key aspects of how school systems function, to understand how qualified human capital is distributed, and to examine how educational inequalities are produced or reduced. International research has long studied teacher mobility, particularly focusing on turnover and transfers between schools, highlighting their implications for the organisational stability of schools and for student learning outcomes (Ingersoll, 2001; Guarino et al., 2006). Several studies show that teacher exits and replacements tend to weaken instructional continuity and reduce student performance, thereby increasing instability within educational contexts (Ronfeldt et al., 2013; Gibbons et al., 2021). Moreover, mobility is not evenly distributed: schools operating in socially disadvantaged areas, for instance, experience higher levels of turnover and greater difficulty in retaining experienced staff, with effects that further amplify existing inequalities (Allensworth et al., 2009; Simon & Johnson, 2015).

Within this line of research, a substantial body of literature has analysed the factors that drive teachers' mobility decisions. Individual preferences and geographical constraints play a key role: several studies show that teachers tend to choose schools located close to their places of origin or training, following a logic of proximity that directly affects recruitment and placement processes (Boyd et al., 2005; Reininger, 2012). This dynamic contributes to creating territorial imbalances in the teacher labour market: some areas struggle to attract teachers, while others are able to ensure a more stable and qualified workforce (Goldhaber et al., 2015). Mobility, therefore, is not a neutral phenomenon, but one that reflects and reproduces pre–existing territorial inequalities within education systems.

The Italian context presents specific dynamics in relation to teacher territorial mobility, closely linked to persistent North–South disparities in the distribution of educational and occupational opportunities, as well as to broader asymmetries characterising the public sector labour market (Rizzica, 2021). In this framework, the literature highlights how the Italian teacher labour market is marked by a structural mismatch between the demand for and supply of teachers, which is embedded in the broader phenomenon of highly skilled interregional mobility (Barbieri et al., 2011; Buonomo et al., 2017).

Teacher geographical mobility represents a historical and structural feature of the Italian school system, shaped by recruitment processes, employment precarity, and persistent territorial inequalities that structure teachers' professional trajectories (Colucci & Gallo, 2017). Changes in initial teacher education systems and in access routes to the profession have also contributed to shaping forms of professional and territorial mobility characterised by institutional fragmentation

and employment instability (Magni, 2019). In this context, a substantial share of teachers originates from Southern regions, while a large proportion of available positions is concentrated in Central and Northern Italy, generating migratory flows that are part of the broader pattern of internal migration of graduates from the South towards the country's more economically dynamic areas (Iammarino & Marinelli, 2015). These dynamics are also influenced by recent changes in recruitment policies and teacher mobility regulations (Desideri, 2025).

Despite the extensive attention devoted to teacher mobility, most studies focus on single episodes of movement or on discrete decisions of entry into and exit from the school system. Much less is known, instead, about life–course territorial trajectories, which simultaneously include places of education and those of professional entry and development. This study aims to address this gap by analysing the geographical trajectories of Italian teachers, reconstructed through four key milestones: the location of secondary school graduation, the location of university graduation, previous teaching experiences, and current school assignment. The study therefore examines different educational and professional pathways of teachers within the Italian education system, with particular attention to their territorial dimensions.

Data Used

This contribution uses data from the Quarta Indagine sui Docenti Italiani, a national survey administered to a representative sample of 9,671 teachers. The survey collects detailed information on teachers' educational and professional trajectories, including where they completed upper secondary education, which is used as a proxy for the respondent's place of origin, as well as their university education, previous teaching experiences, and current teaching position. These data make it possible to reconstruct teachers' territorial mobility trajectories across different stages of their educational and professional lives.

Methodology

Drawing on these data, a typology of teachers' territorial mobility trajectories is developed by integrating mobility during the educational phase, mobility throughout the professional career, and the geographical location of the current teaching position. All these geographical measures are harmonized at the same level of detail, namely the regional groupings commonly used for statistical purposes: North–East, North–West, Centre, South, and Islands. Given this level of detail, mobility is identified whenever the respondent reports having moved outside the regional grouping in which they completed secondary school, studied at university or has previously taught, depending on the specific transition. The resulting classification identifies six profiles: teachers with no territorial mobility (62.8%); return trajectories following educational mobility (7.7%) or work–related mobility (10.6%); trajectories characterised by the absence of return to the area of origin after educational mobility (4.8%) or work–related mobility (10.3%); and trajectories involving multiple territorial moves throughout the professional career (3.9%). The socio–demographic and professional composition of these different profiles is then examined through both bivariate and multivariate analyses.

Results and discussion

The analysis of territorial trajectories shows a prevalence of teachers who did not move either for education or for work, indicating a substantial stability of the teaching workforce and suggesting that, for many, the choice of profession is accompanied by a strong degree of territorial continuity. However, this stability does not exhaust the picture: trajectories involving mobility – whether as returns to places of origin or as continued residence away from them – are far from marginal and are unevenly distributed across subgroups. In Southern regions, return trajectories are more frequent, whereas in Northern regions trajectories that do not involve a return to places of origin are more common, a pattern consistent with persistent territorial differences in employment opportunities and in the capacity of local contexts to attract and retain qualified personnel. Differences by gender, school level, and disciplinary area are limited, but some patterns emerge: women display greater territorial continuity, while more complex trajectories are observed in technical and vocational

schools, in line with the specific features of recruitment processes and the organisational structure of these segments of the school system. Years of service also introduce some differentiation: more experienced teachers tend to exhibit more stable trajectories, whereas intermediate career stages show greater diversification, suggesting that mobility is particularly characteristic of the central phases of the teaching career. Exploratory analyses further indicate that these territorial and professional differences are associated with different motivations for entering the teaching profession and with varying levels of job satisfaction in carrying out the teaching role.

Conclusions

The findings highlight the importance of treating teachers' territorial mobility as a heterogeneous and life–course process, unfolding across both educational and professional stages. The different trajectories identified in this study show that mobility is unevenly distributed and closely intertwined with territorial and occupational differences, reflecting the structure of the teacher labour market and the organisation of the school system. This heterogeneity suggests that different territorial trajectories may be associated with different ways of experiencing and enacting the teaching profession and, consequently, may have implications not only for teachers' well–being but also for students' learning processes.

Keywords: Teacher mobility; Territorial trajectories; Italian school system; Teacher labour market

Measuring Teacher Self–Efficacy Across School Levels: Validation of a Multidimensional Scale in the Italian Context

Paolo Guandalini – Iacopo Moreschini

Introduction

In recent decades, the teaching profession has become progressively more complex. The increase in special educational needs, the growing heterogeneity of classrooms, the presence of both explicit and implicit school dropout, and the overlap of instructional, educational, organizational, and interpersonal responsibilities have significantly expanded the scope of teachers' work. Added to this transformation is a diversity within the teaching profession that extends not only to teachers' individual characteristics but also to the grade levels at which they work, with significant differences in their responsibilities, organizational contexts, relationships with students and families, and professional expectations.

In this context, teachers' self–efficacy represents a crucial resource. The perception of being able to influence learning processes, manage the classroom, engage students, and employ effective teaching strategies is not merely an individual trait but an indicator of the perceived professional capacity with which teachers navigate increasingly diverse contexts (Bandura, 1977; Tschannen–Moran & Hoy, 2001). The literature has shown that teachers' self–efficacy is associated, on the one hand, with higher levels of job satisfaction, motivation, and retention in the profession, as well as a lower risk of burnout and work–related stress (Brouwers and Tomic, 2000; Canrinus et al., 2012; Caprara et al., 2006; Chesnut and Burley, 2015; Skaalvik and Skaalvik, 2007); and, on the other hand, with teachers' greater ability to influence student learning and motivation (Holzberger, Philipp, and Kunter, 2013; Lauermann and ten Hagen, 2021).

However, in the Italian context, the measurement of teachers' self–efficacy still has significant limitations. The available scales have been validated primarily in the context of international surveys such as OECD TALIS, which include a limited subset of teachers, predominantly from lower secondary schools (OECD, 2025). This results in only partial coverage of the Italian teaching workforce, leaving other levels of education largely uncovered.

International surveys remain essential for placing Italian teachers within a cross–national comparison, but homogeneous samples such as those used in TALIS allow only a limited view of how self–efficacy is distributed within a diverse teaching workforce, treating as uniform a resource that, on the contrary, can vary significantly depending on work contexts, the responsibilities assigned to teachers, and the organizational conditions in which they operate.

Given this limitation, this paper proposes the validation of a teacher self–efficacy scale based on data from a recent national survey involving 283 schools at all levels. The scale incorporates three dimensions well–established in the international literature: the ability to engage students, classroom management, and the application of effective teaching strategies. In addition, the proposed scale introduces a fourth dimension designed to capture teachers' perceptions of their ability to influence their work environment. In this way, this study aims not only to provide a measurement tool suited to the Italian context and applicable to different levels of education, but also to highlight a component of self–efficacy that is particularly relevant in a profession shaped by organizational, educational, and institutional demands.

The study pursues two main objectives: first, to assess the factorial structure and reliability of a multidimensional teacher self–efficacy scale applicable across different levels of the Italian school system; second, to examine whether the resulting scores vary by school level and selected individual and professional characteristics. The analyses of group differences are therefore intended as a substantive application of the validated measurement model.

Data

The analysis is based on a sample of 9,671 teachers surveyed as part of the Fourth National Survey on Teachers using a two–stage sampling design. The primary sampling units consist of 283 schools, randomly selected from the population of Italian schools, within which the teachers included in the survey were subsequently identified and interviewed during the spring of 2025 (Argentin, 2026). When adjusted to the overall teacher population, 36.9% of the teachers interviewed were elementary school teachers and 25% were lower secondary school teachers, while the remainder were divided among high school teachers (15.9%), technical school teachers (15.4%), and vocational school teachers (6.8%). The analysis is based on a battery of 20 four–point Likert–scale items (ranging from "Not at all" to "Very much") related to teachers' perceived self–efficacy. The battery is organized into four dimensions, three of which – Student Engagement (6 items), Instructional Strategies (5 items), Classroom Management (5 items) – reflect the items present in the TALIS survey, while the fourth – Influencing the school context (4 items) – is an elaboration by the authors designed to capture an additional dimension of teachers' professional self–efficacy.

Methodology

In order to distinguish the exploratory phase from the confirmatory phase, the analytical sample was randomly divided into two subsamples of equal size using a 50/50 split–sample procedure, which was carried out prior to the estimation of the factor models. The first subsample was used for exploratory factor analysis (EFA), and the second for confirmatory factor analysis (CFA), in order to test the theoretical structure of the battery on independent data and reduce the risk of overfitting the model to the data. All analyses were conducted taking sample weights into account and performed in R using the psych and lavaan packages.

Given the ordinal nature of the items, the correlation matrices were estimated using polychoric correlations.

Exploratory factor analysis was conducted using principal axis factoring, resulting in a four–factor solution consistent with the theoretical structure of the battery; since the dimensions were expected to be correlated, an oblique rotation (oblimin) was applied.

Two confirmatory factor analyses were subsequently conducted on the second subsample. The first CFA specified a theoretical structure of four correlated latent factors: ability to engage students, application of effective teaching strategies, classroom management, and ability to influence the school environment. The second CFA estimated a second–order hierarchical model, based on the

assumption that the four first–order dimensions are not independent but represent components subordinate to a single general factor of teacher self–efficacy, which accounts for their covariation. The final model adopted is the second–order CFA model.

Results and Discussion

Initial Findings show that internal consistency was good to excellent across all subscales, with ordinal alpha and omega coefficients ranging from .74 (Influence on Contexts) to .89 (Student Engagement), and reaching .95–.96 for the General Self–Efficacy scale. The hierarchical omega (.83) indicates that a substantial proportion of the reliable variance in the battery is attributable to the general self–efficacy factor rather than to the specific dimensions, supporting the use of a composite score alongside the four subscale scores.

The EFA conducted on the first subsample confirmed the plausibility of a four–factor structure (RMSR = .02, TLI = .955, RMSEA = .053; 58% of variance explained), with moderate correlations between factors (.49–.66). However, the empirical structure that emerged only partially overlaps with the theoretical one: some items exhibit cross–loadings on multiple factors, suggesting a partial overlap between the dimensions.

The two confirmatory models show substantially equivalent fit to the data. The scaled indices indicate a satisfactory fit (CFI/TLI > .95, RMSEA = .060, SRMR = .040), while the robust indices paint a more cautious picture (CFI/TLI ≈ .90–.91, RMSEA ≈ .080), suggesting an acceptable but not optimal fit.

The most significant finding is that the second–order model, which imposes the constraint that the correlations among the four dimensions are fully explained by a general superordinate factor, does not result in any loss of fit compared to the model with four correlated factors. This result, together with the high latent correlations among the factors (.831–.945) and the high proportion of common variance explained by the general factor (ECV = .74), further supports the idea that the four dimensions are not independent but rather expressions of a single construct of teacher self–efficacy, justifying the preference for the hierarchical model due to its greater parsimony.

Standardized scores for teacher self–efficacy vary significantly by grade level, both for the general factor and for all four dimensions (ANOVA, $p < .001$). The pattern is consistent and monotonically increasing across all dimensions: elementary school teachers report the highest levels of self–efficacy, followed by lower secondary school teachers (close to the overall mean), while upper secondary school teachers show the lowest values.

Overall, the results indicate that the association between individual and professional characteristics and self–efficacy varies significantly across the different school levels, with a more favorable profile in elementary school and more widespread challenges in high school. This evidence suggests that teachers' self–efficacy cannot be viewed as a stable, context–independent individual trait, but must be interpreted in relation to the organizational, instructional, and relational conditions specific to the various segments of the school system. The more problematic picture observed in upper secondary school reflects the greater complexity of teaching work at this level, where subject specialization, the fragmentation of curricula, and the management of adolescent students can make it more difficult for teachers to perceive themselves as effective both in their teaching and in fostering an inclusive, learning–friendly school environment.

At the same time, differences in self–efficacy related to teachers' individual features vary within school levels. Gender differences are limited to specific dimensions such as influence within the classroom or the ability to engage students, while age differences emerge primarily in lower secondary school, where teachers under 40 show lower levels than their older colleagues. This finding can be cautiously interpreted as a possible effect of accumulated professional experience, but it does not allow for a clear distinction between the effects of age, length of service, and professional position. Overall, the results therefore suggest that self–efficacy should be considered a situated construct, sensitive not only to teachers' individual characteristics but also to the constraints and resources of the school contexts in which they work.

Conclusions

This study contributes to the measurement of teacher self–efficacy in the Italian context by validating a multidimensional scale based on a large national sample of teachers across different school levels. The results support the psychometric robustness of the instrument: internal consistency is good to excellent across the four dimensions, the exploratory analyses confirm the plausibility of the expected factor structure, and the confirmatory analyses show that the data are adequately represented by a hierarchical model in which the four dimensions are expressions of a broader general factor of teacher self–efficacy.

The validation results support the use of both a general teacher self–efficacy score and four dimension–specific scores. The strong general factor confirms the coherence of the construct, while the four subscales allow researchers to distinguish between self–efficacy in student engagement, instructional strategies, classroom management, and influence on the school context.

A further contribution of the study is the inclusion of a dimension concerning teachers' perceived ability to influence the school context. This component extends the traditional measurement of teacher self–efficacy beyond classroom–level activities and recognizes that teachers' professional agency also operates within organizational and institutional settings. The results suggest that this dimension can be integrated into a broader self–efficacy framework without weakening the coherence of the scale.

The study provides baseline evidence on the distribution of teacher self–efficacy across different levels of the Italian school system. This is a relevant contribution, since comparative international surveys have generally offered only partial coverage of the teaching workforce, focusing mainly on lower secondary education. By including elementary, lower secondary, and upper secondary teachers within the same measurement framework, the study makes it possible to observe differences that would otherwise remain hidden in more restricted survey designs.

The results show a clear gradient by school level: elementary school teachers report the highest levels of perceived self–efficacy, lower secondary teachers occupy an intermediate position, and upper secondary teachers show the lowest scores across the general scale and the four specific dimensions. These findings suggest that teacher self–efficacy is strongly shaped by the institutional and organizational contexts in which teachers work. At the same time, the validated scale provides a common metric through which these differences can be described, monitored, and further investigated in future research.

The analysis of differences by school level and teacher characteristics should therefore be read as an application of the validated measurement model. These results show that self–efficacy is not distributed uniformly across the teaching workforce, but varies according to the school contexts in which teachers work. However, the main contribution of the paper remains methodological: the study provides evidence for a reliable and theoretically grounded instrument that can be used to investigate teacher self–efficacy across the Italian school system.

Keywords: Teacher self–efficacy; Scale validation; Educational levels; Italian teachers

Teacher Voice, Student Engagement and School Effectiveness: Evidence from Two Exploratory Surveys

Sebastiana Fisicaro

Introduction

The reduction of educational inequalities and school dropout has become a strategic priority of the Italian educational system, particularly within the framework of the National Recovery and

Resilience Plan (PNRR). Recent evidence provided by the Italian National Institute for the Evaluation of the Education System (INVALSI) continues to highlight persistent challenges related to implicit school dispersion, unequal learning outcomes, and territorial disparities, pointing to the need for interventions capable of promoting both academic achievement and meaningful student participation. Within this context, student engagement has emerged as a key factor in preventing educational disengagement and supporting school success. International research consistently identifies engagement as a multidimensional construct associated with academic achievement, persistence, well–being, and educational attainment (Fredricks et al., 2004; Kahu, 2013). However, engagement is not solely a student attribute. It develops through the interaction between learners, teaching practices, classroom relationships, and school environments, making teachers central actors in its construction. Teachers occupy a strategic position in educational improvement processes. Through their daily professional decisions, they influence not only learning opportunities but also students' motivation, participation, and sense of belonging. For this reason, understanding how teachers interpret effective teaching, student engagement, and educational innovation is essential for understanding how improvement policies are translated into practice. Research on educational change has repeatedly shown that reforms are unlikely to generate sustainable effects unless teachers actively participate in interpreting, adapting, and enacting innovation within their professional contexts (Priestley et al., 2015; Fullan, 2020). The present study is situated within a broader programme of professional development and school support promoted by Sophia 3.0, a professional network that has been supporting teachers and school leaders through training activities, communities of practice, and innovation projects for more than a decade. Two exploratory surveys were developed and administered to teachers working in schools engaged in educational innovation processes, including PNRR–related initiatives, particularly in contexts characterised by educational vulnerability and risks of school disengagement. Rather than measuring student engagement directly, the study investigates teachers' professional representations of the conditions that make learning possible and support educational success. By bringing teacher voice to the centre of the analysis, the research contributes to current debates on school effectiveness and school improvement, highlighting professional knowledge as a valuable source of evidence for understanding educational change. From a school effectiveness perspective, understanding how teachers interpret the conditions that support engagement and learning is essential for identifying the factors that contribute to educational quality and improvement.

The study addresses the following research questions:

RQ1. How do teachers define effective teaching practices?

RQ2. Which conditions do teachers identify as most relevant for supporting student learning?

RQ3. How do teachers conceptualise student engagement and the strategies required to foster it?

RQ4. What implications do teachers' perspectives have for school effectiveness and school improvement?

Data Used

The study draws on data collected through two complementary online surveys administered between March and April 2026 within the Sophia 3.0 network. Sophia 3.0 is a professional network and cultural association that has supported teachers, school leaders, and educational innovation initiatives for more than a decade through training programmes, communities of practice, research activities, and school improvement projects. The surveys were developed within this context as part of a broader effort to understand teachers' professional experiences, perceptions, and needs in schools engaged in educational innovation processes, including those supported by the Italian National Recovery and Resilience Plan (PNRR). The decision to employ two separate instruments was motivated by the intention to explore two interconnected dimensions of educational practice. The first survey investigated teachers' professional development needs, perceptions of educational innovation, and views on the challenges currently affecting teaching and learning. A total of 70 teachers from different school levels and professional roles participated in the survey. Participants

were invited to reflect on issues related to effective teaching, student motivation, inclusion, educational innovation, professional learning, and the transformation of learning environments. The second survey focused specifically on student engagement and involved 14 teachers representing different educational sectors. The questionnaire explored teachers' perceptions of the factors that facilitate or hinder student participation, the strategies used to promote engagement, and the pedagogical approaches considered most effective in supporting meaningful learning experiences. Particular attention was devoted to the relationship between engagement, teaching practices, classroom climate, and educational success. Both questionnaires combined closed–ended and open–ended questions, allowing participants not only to express preferences and priorities but also to articulate the reasoning underlying their professional choices and experiences. The instruments were collaboratively designed and refined within the Sophia 3.0 network through discussions involving researchers, school leaders, and teachers, ensuring their relevance to current educational challenges and professional practice. Participation in both surveys was voluntary and anonymous. This methodological choice was intended to encourage the free expression of professional perspectives and to minimise potential organisational influences on participants' responses. The resulting dataset provides access not only to teachers' opinions but also to the meanings, experiences, and professional representations through which educators interpret effective teaching, student engagement, educational innovation, and the conditions that support learning and school improvement in contemporary educational contexts.

Methodology

This study adopts an exploratory mixed–methods design aimed at investigating teachers' professional perspectives on effective teaching, student engagement, and school improvement within the context of ongoing educational innovation processes. The research combines quantitative and qualitative evidence collected through two online surveys administered within the Sophia 3.0 professional network. The choice of a mixed–methods approach was motivated by the need to capture both recurring patterns in teachers' responses and the meanings that educators attribute to their professional experiences. The first survey explored teachers' professional development needs, perceptions of educational innovation, and views on the challenges currently affecting teaching and learning. The second focused on student engagement, examining perceived barriers to participation, strategies adopted to foster involvement, and teachers' evaluations of effective pedagogical approaches. Participants represented different school levels, professional roles, and years of teaching experience, providing a diverse range of perspectives on contemporary educational practice. Both instruments included closed–ended and open–ended questions. Quantitative data were analysed through descriptive statistics in order to identify recurrent trends, priorities, and areas of concern. Qualitative responses were examined through an inductive thematic analysis, allowing categories and themes to emerge directly from teachers' narratives rather than being imposed a priori. Particular attention was devoted to identifying recurring representations of effective teaching, student engagement, educational innovation, and school improvement.The integration of quantitative and qualitative evidence enabled a more comprehensive interpretation of teachers' professional perspectives and facilitated the triangulation of findings across the two surveys. Rather than measuring student engagement directly, the study investigates how teachers interpret the conditions that facilitate or hinder participation, learning, and educational success. Consistent with current research on teacher agency and school improvement, teacher voice is considered a valuable source of evidence for understanding how educational policies and innovation initiatives are experienced, interpreted, and enacted within school contexts. The methodological focus on professional representations therefore provides insight not only into teachers' perceptions but also into the processes through which educational change is translated into everyday practice.

Results and Discussion

The findings provide a coherent picture of how teachers interpret the conditions that support learning, student participation, and school effectiveness. Across different school levels and

professional roles, respondents consistently identified student engagement as one of the most significant challenges of contemporary teaching. Difficulties were mainly associated with declining attention, limited motivation, increasingly heterogeneous classrooms, and the growing influence of digital distractions. Teachers frequently linked student engagement to teaching practices, classroom relationships, and instructional choices, suggesting that participation and motivation are strongly influenced by the quality of the learning environment. A recurring theme emerging from both surveys concerns the perceived limitations of predominantly transmissive approaches. While respondents did not reject direct instruction itself, they frequently associated low levels of participation with frontal and passive forms of teaching. By contrast, laboratory activities, cooperative learning, guided discussion, tutoring practices, problem–based learning, gamification, and the meaningful integration of digital technologies were consistently identified as effective strategies for fostering motivation, participation, and active learning. These responses indicate a widespread orientation towards learner–centred pedagogies and participatory teaching practices. The analysis also highlights a strong relationship between student engagement and teachers' professional learning needs. Respondents repeatedly emphasised the importance of practical training opportunities, operational tools, examples of effective practices, peer exchange, and time for experimentation. Professional development was considered particularly valuable when directly connected to classroom challenges and capable of supporting informed pedagogical decision–making. The preference for workshops, collaborative learning opportunities, and blended professional development pathways further confirms the importance attributed to practice–oriented forms of professional learning. Teachers also tended to interpret educational innovation primarily in terms of changes in teaching practices rather than through organisational or infrastructural transformation alone. Although many respondents recognised the relevance of recent innovation initiatives, including those promoted through the PNRR framework, they consistently emphasised methodological innovation, inclusion, classroom management, student motivation, and the integration of digital technologies as the most critical dimensions of school improvement. These findings suggest that sustainable educational change depends largely on teachers' professional agency and on their capacity to interpret, adapt, and translate policy priorities into meaningful learning experiences. Overall, the study positions teacher voice as a valuable source of evidence for understanding school effectiveness and educational improvement. Engagement, innovation, and school development emerge not as separate dimensions but as interconnected aspects of professional practice, highlighting the central role of teachers in shaping educational quality and fostering meaningful learning opportunities. These findings acquire particular significance in the context of schools involved in PNRR initiatives aimed at reducing educational inequalities and preventing school disengagement. In line with recent INVALSI evidence on persistent risks of implicit school dispersion, the study suggests that strengthening student engagement requires not only organisational and structural investments but also sustained support for teachers' professional learning, agency, and pedagogical innovation.

Conclusions

The study provides evidence that teachers perceive student engagement as a critical condition for learning, educational success, and school improvement. Across different school levels and professional roles, participants identified declining attention, reduced motivation, classroom heterogeneity, and digital distractions as major challenges requiring innovative pedagogical responses. These findings are particularly relevant in the context of schools involved in PNRR initiatives, where reducing educational inequalities and preventing school disengagement represent strategic priorities. The results reveal a strong convergence around active, participatory, and learner–centred approaches. Teachers consistently recognised laboratory activities, cooperative learning, guided discussion, problem–based learning, and the purposeful integration of digital technologies as effective strategies for promoting meaningful participation and learning. At the same time, they expressed a clear demand for practice–oriented professional development, collaborative learning opportunities, and concrete tools capable of supporting everyday teaching practice. A

significant finding concerns the way teachers interpret educational innovation. Rather than associating improvement primarily with organisational reforms or infrastructural investments, participants emphasised the importance of transforming pedagogical practices and strengthening professional competences. This highlights the central role of teacher agency in translating policy priorities and innovation initiatives into meaningful educational experiences for students. From a broader perspective, the findings support the view that teacher voice constitutes a valuable source of evidence for understanding school effectiveness and informing improvement strategies. In a national context where INVALSI data continue to draw attention to persistent risks of educational vulnerability and implicit school dispersion, listening to teachers' professional perspectives can contribute to identifying context–sensitive and sustainable pathways for educational change. Future research should extend the investigation to larger and more diverse samples and explore the relationship between teachers' perceptions, students' experiences, and educational outcomes. Nevertheless, the present study suggests that policies aimed at reducing educational inequalities and preventing school failure cannot rely exclusively on structural investments or organisational reforms. Their effectiveness ultimately depends on teachers' capacity to create engaging, inclusive, and meaningful learning environments capable of supporting participation, achievement, and long–term educational success.

Keywords: Teacher Voice; Student Engagement; Educational Innovation; School Effectiveness

Leading Through Technology: The Role of School Principals' Digital Leadership in Shaping Educational Efficiency

Mara Soncin – Tommaso Agasisti

Introduction

While digital technologies are increasingly embedded in European education systems, their role in improving school efficiency—that is, how well schools transform inputs into educational outcomes—remains underexamined. A growing body of research highlights the potential of digital tools to support teaching and learning (Tamim et al., 2011; Escueta et al., 2020), but fewer studies have assessed how such tools, when integrated strategically, contribute to the overall productivity of schools (Biagi & Loi, 2013; Bakker, 2021). In parallel, the literature on school leadership has consistently shown that principals play a central role in shaping school effectiveness through goal–setting, instructional coordination, and professional development (Leithwood et al., 2004; Hallinger, 2011; Robinson et al., 2008). However, the intersection between school leadership and digitalization – what is increasingly termed digital educational leadership – remains relatively underexplored, particularly in large–scale, efficiency–focused research (Dexter, 2011; Hamzah et al., 2021). While there is some evidence that digital leadership fosters innovation and teacher engagement (Saeed & Kang, 2024), we still know little about how different aspects of digital leadership interact and influence the ability of schools to convert both traditional and digital resources into student outcomes. This paper addresses this gap by investigating the relationship between three distinct components of digital leadership and school efficiency across the countries of the European Union, using microdata from OECD PISA 2022. Rather than treating digital leadership as a single latent construct, we operationalize it through three complementary dimensions derived from the PISA 2022 school questionnaire:

1. Digital policies, capturing the existence of school policies, guidelines, and programmes regulating and promoting the educational use of digital technologies;
2. Teacher collaboration, reflecting the extent to which schools foster collaboration among teachers in the use and development of digital teaching practices;

3. Teacher skills, measuring principals' perceptions of teachers' digital and pedagogical competences, as well as the time and opportunities available for developing these skills.

The goal is twofold: (1) to assess the individual contribution of each leadership dimension to school efficiency, and (2) to explore their interactions, i.e., whether certain combinations of leadership practices amplify or diminish their overall effect.

This study offers an innovative contribution to the field of education research by integrating advanced efficiency analysis with a multidimensional operationalization of digital leadership, based on large–scale, cross–national data. Unlike most prior studies, which either focus on the availability of ICT or treat leadership as a single latent factor, this work disentangles the effects of three distinct leadership dimensions – digital policies, teacher collaboration and teacher skills – and examines both their individual and interactive impacts on school efficiency. Methodologically, the use of sophisticated efficiency models represents a novel approach in educational settings, allowing for a more reliable estimation of how school–level practices shift the education production frontier.

Data and Methodology

From a methodological standpoint, this study adopts a robust non–parametric efficiency estimation strategy, applying both conditional DEA (Cazals, Florens & Simar, 2002) and Robust Order–m DEA model (Daraio & Simar, 2007). The latter is particularly suited to capturing efficiency under conditions of environmental heterogeneity and data uncertainty. The conditional analysis allows for the inclusion of contextual factors that may condition efficiency, which in this case will be the indicators of digital leadership. By comparing results from both approaches, we assess the consistency and robustness of the estimated relationship between school leadership and educational efficiency. This dual–model strategy enables us to test whether digital leadership practices systematically shift the educational production frontier, rather than merely affecting schools' position relative to it. Inputs include resources such as student–teacher ratio, richness of educational resources and school socio–economic status. Outputs are the school–level average scores in reading, mathematics, and science from the PISA 2022 assessments. Additionally, interaction terms between the three indicators are included to capture complementarity or substitution effects: for example, whether infrastructure leadership is more effective when combined with strong instructional coordination.

Results and Discussion

Preliminary findings indicate that school efficiency differs substantially across European countries, with average global efficiency around 0.85 and higher within–country (local) efficiency (about 0.93), suggesting that part of the observed differences is driven by national contexts. Accounting for digital instructional leadership increases estimated conditional efficiency, confirming that leadership practices shape the educational production process. Among the examined dimensions, the existence of clear digital policies and guidelines shows the strongest association with efficiency, particularly when combined with teachers' digital skills and adequate ICT support. When digital intensity is explicitly included among the production inputs, ICT support becomes the most influential leadership component, reinforcing the importance of coupling technological investments with effective school leadership and teacher capacity building.

The findings provide consistent evidence that digital instructional leadership is associated with higher levels of school efficiency across European education systems. While substantial cross–country heterogeneity emerges in unconditional efficiency scores, accounting for digital leadership systematically improves the estimation of school performance, suggesting that leadership practices constitute an important contextual factor in the educational production process. In particular, the existence of clear school–wide digital policies appears to be the most influential dimension of digital leadership. This effect is further strengthened when coherent digital policies are combined with teachers' digital skills and adequate ICT support, highlighting the complementary nature of these leadership practices.

These findings contribute to the literature in two main ways. First, they extend previous research on school efficiency by demonstrating that managerial and leadership practices should be considered alongside traditional school inputs when analysing educational production. Second, they enrich the emerging literature on digital leadership by showing that its contribution extends beyond the adoption of technology itself and is instead rooted in the principal's capacity to create an organisational environment that supports the effective pedagogical use of digital resources.
From a policy perspective, the results suggest that investments in school digitalisation should not focus exclusively on technological infrastructure. Equally important are leadership development programmes that enable school principals to design and implement coherent digital policies, foster collaboration among teachers, strengthen teachers' digital competences, and ensure adequate technical and pedagogical support within schools. Policies targeting these complementary dimensions are likely to produce greater improvements in school efficiency than isolated investments in digital equipment alone.

Keywords: Efficiency; Digital leadership; School principals; International data

Better Environment, Better Teaching: The Power of Teacher Networks

Teodora Uberti – Giulietta Zanga

Introduction and background

A growing body of research shows that the way a school is organized — the density and quality of exchange among its members, the degree of hierarchy, and the extent to which different organizational units cooperate — is a key determinant of teachers' professional experience. Positive and functional collaboration among teachers improves classroom practice and is systematically associated with higher job satisfaction and well–being (Aquario, Ghedini & Bresciani Pocaterra, 2017), with benefits that extend to students in terms of participation, motivation, and the development of social skills. Teachers who are more centrally connected within professional networks tend to feel more confident and satisfied, benefiting from greater opportunities for exchange, feedback, and support (Uberti & Zanga, 2026), and the students of well–integrated teachers often obtain better academic outcomes (Moolenaar, 2012; Siciliano, 2017; Daly, Moolenaar, Bolivar & Burke, 2010).
While this literature examined collaboration from the standpoint of the individual teacher and their ties to colleagues, here we focus on the perception of teachers of the overall school organisation , i.e. teachers' perception of the school as a whole, its overall organizational structure, the degree of hierarchy, and the extent of cooperation between organizational levels and groups. This macro–level perception is not incidental to the school's functioning but is also the outcome of leadership choices. Principals exercise a measurable, though largely indirect, influence on school effectiveness through the goals, structures and routines they establish (Hallinger & Heck, 1998), and effective leadership is increasingly understood as distributed across formal and informal roles rather than concentrated in a single office (Spillane, 2006). Leadership practices that foster shared goals, professional community and collaborative structures are the school–level factor most strongly associated with improved outcomes (Leithwood, Louis, Anderson & Wahlstrom, 2004), and the relational trust that binds principals, teachers and staff together is itself a core resource for school improvement (Bryk & Schneider, 2002).

Aim and research questions

For the first time in Italy, a large–scale survey explored how teachers perceive the organization of, and their own position within, their school's professional network. The study addresses the following

reaserch question: how is teachers' job satisfaction linked to their perception of the school's organizational structure and of professional relationships with colleagues?

Data and methods

The data derive from "IV Insegnanti" survey, funded by the Fondazione Bolton, which interviewed nearly 10,000 Italian teachers. A random subsample of 4,822 respondents was asked a dedicated battery of questions distinguishing three complementary levels of analysis: their perception of the school's organizational structure (i.e. macro level: five categories ranging from hierarchical organizations with limited inter–level exchange to fully integrated, non–hierarchical networks, or integrated structures composed of cooperating versus non–cooperating groups); their perceived position within the network of colleagues (i.e micro level: coordinators, collaborators, marginal teachers, and marginal brokers, distinguished by level of coordination); and the presence and characteristics of a trusted colleague, with whom professional problems and opinions are shared (i.e. ego–alter level). Job satisfaction is measured on an 11–point scale (0 = "not at all satisfied" to 10 = "completely satisfied") in response to the question "Overall, how satisfied are you with your current job?". The association between these perceptions and job satisfaction is estimated through OLS regression with robust standard errors (N = 4,814; R^2 = 0.113), controlling for socio–demographic characteristics (gender, age, caregiving responsibilities), professional characteristics (career mobility, union membership, prior work experience, current and past roles, motivation, perceived job security, subject area), and school–level characteristics (geographical macro–area, infrastructural conditions, and availability of facilities).

Results

The central result of the study concerns the macro–level perception of school organization. Teachers who perceive their school as an integrated, non–hierarchical network, or as organized in cooperating groups, report the highest levels of job satisfaction, while a hierarchical structure with limited exchange between organizational levels — the modal perception, reported by 24% of teachers — is associated with the lowest satisfaction among the organizational categories. Compared with this hierarchical–with–limited–exchange baseline, all more integrated or exchange–intensive configurations are associated with significantly higher satisfaction: integrated organizations structured around cooperating groups show the largest coefficient ($b = 0.356$, $p < 0.01$), followed by a fully integrated, non–hierarchical network ($b = 0.305$, $p < 0.01$) and a hierarchical structure with intense inter–level exchange ($b = 0.265$, $p < 0.01$). Notably, even an integrated structure whose constituent groups do not cooperate with one another retains a positive, if smaller, association with satisfaction ($b = 0.132$, $p < 0.05$), suggesting that integration itself carries an independent benefit beyond cooperation. These patterns are consistent across school levels (primary, lower– and upper–secondary), with upper–secondary teachers somewhat more likely to report hierarchical perceptions and somewhat less satisfied overall ($b = –0.095$, $p < 0.01$, relative to primary teachers).

The remaining two levels of analysis reinforce and help interpret the organizational–level finding rather than compete with it. At the micro level, teachers who see themselves as part of cohesive, collaborative groups — particularly coordinators embedded in a collaborative context — report the highest satisfaction (b up to 0.479, $p < 0.01$), whereas isolated teachers report the lowest, and teachers acting as marginal brokers between groups (16% of the sample) show lower–than–average satisfaction, consistent with the notion that bridging disconnected parts of a fragmented organization is professionally costly. At the ego–alter level, having no trusted colleague to turn to is associated with markedly lower satisfaction ($b = –0.283$, $p < 0.01$; mean satisfaction 6.9 versus 7.7 among teachers with a trusted colleague), while self–efficacy shows the single largest association with job satisfaction in the model ($b = 0.238$, $p < 0.01$). Taken together, the three levels indicate that the same organizational logic — integration over fragmentation, cooperation over isolation — operates from the whole–school structure down to the individual teacher's closest professional tie.

Discussion and implications

These findings speak directly to the question of what makes a school work, by showing that teachers' satisfaction is shaped first and foremost by how the school as a whole is organized — not only by individual relationships. Because organizational structures of the kind captured here (hierarchy, inter–level exchange, inter–group cooperation) are largely the product of principals' choices about governance, communication routines and shared decision–making, our results provide indirect but consistent survey evidence for the claim that school leadership operates on teacher outcomes primarily through the organizational and relational conditions it creates (Hallinger & Heck, 1998; Leithwood et al., 2004; Spillane, 2006), and that relational trust between principals, middle–level coordinators and teachers is a tangible resource rather than a soft or secondary concern (Bryk & Schneider, 2002).

For school principals and policy–makers, the practical implication is that investments in integrated, cooperative organizational structures — cross–level committees, cooperating teacher groups, and distributed coordination roles — are likely to pay off in terms of teacher satisfaction and, through it, in classroom quality and student outcomes.Keywords: CBT, impact, drag and drop, dropdown menus

Keywords: Italian teachers; Networks perception; Job satisfaction; Integrated organization

SESSION 3. INTERNATIONAL LARGE–SCALE ASSESSMENTS (ILSAS)
METHODS AND RESULTS
ORGANIZER: UNIVERSITY OF BATH
COORDINATORS: MARIA MAGDALENA ISAC – ANDRES SANDOVAL HERNANDEZ
NOVEMBER 4TH: 2.00 P.M. – 4.00 P.M. {AULA A1.1 – RESEARCH 2}

From Environmental Topics to Classroom Practice: Action–Based Instruction and Environmental Knowledge in TIMSS 2023 Grade 4 Italy

Giada Viotti – Inés García-Bohórquez – Camilo Ruiz – Laura Palmerio

Introduction

Evidence on the environmental consequences of human activity has made it increasingly difficult to separate ecological and climate–related challenges from the ways in which societies organize their re–lationship with the natural world. These challenges cannot be addressed only through technical or regulatory measures, since they also involve educational, social, and cultural dimensions. Education is therefore part of the response, as schools can help students move from understanding environmental issues to recognizing how these issues are reflected in everyday practices and in their own relationship with nature and the environment (Anderson, 2012; Mochizuki & Bryan, 2015). International large–scale assessments have progressively responded to this need by expanding their frameworks beyond traditional cognitive domains and incorporating dimensions related to sustainability, citizenship, and the environment. TIMSS 2023 represents a further step in this direction through the Environmental Attitudes and Behaviors Framework, which expands the environmental knowledge component al–ready included in TIMSS 2019 and links it to students' environmental attitudes, behaviors, and learn–ing contexts documented through the Context Questionnaires (Mullis & Martin, 2017; Yin & Foy, 2020; Reynolds & Komakhidze, 2022). In TIMSS 2023, environmental awareness is conceptualized as a multidimensional construct that includes environmental knowledge, environmental attitudes, and environmental behaviors. Environmental knowledge is assessed through science items related to natu–ral processes, human impacts on the environment, and the use of natural resources. Although TIMSS does not directly observe classroom interaction or the quality of teaching processes, its combination of cognitive, attitudinal, behavioral, and contextual indicators provides a valuable basis for examining how environmental and sustainability education is represented within school systems. By including Grade 4, TIMSS 2023 allows these dimensions to be examined from the primary school years, when environmental knowledge and attitudes are still strongly connected to children's concrete experiences, everyday practices, and familiar places (Green & Somerville, 2015).

Regarding teaching practices, TIMSS 2023 relies on science teachers' reports to describe how envi–ronmental topics are addressed in the classroom. These data allow the analysis to distinguish between practices more closely related to the discussion of environmental issues and practices involving stu–dents in activities, experiences, or actions connected to these topics. This distinction draws on research in environmental and sustainability education, which shows that environmental learning is not limited to the acquisition of scientific information, but also concerns the ways students interpret environmental issues and relate them to possible forms of action (Jensen & Schnack, 1997; Mogensen & Schnack, 2010; Varela–Losada et al., 2016; Sass et al., 2020). For the present analysis, these teacher question–naire items are grouped into two TIMSS–based indicators: discussion–based environmental instruction and action–based environmental instruction. The analysis starts from the

idea that teaching practices in Environmental and Sustainability Education (ESE) can be read at different levels. Although TIMSS provides school–level information on ESE, science teachers' reports refer to practices enacted in spe–cific classroom contexts. For this reason, the study examines whether the association between reported teaching practices and students' environmental knowledge changes when these practices are consid–ered at the class level and at the school level. This distinction is relevant because the aggregation of teaching practices at school level may obscure part of their pedagogical meaning and lead to interpret–ing aggregate associations as if they reflected classroom–level processes (Robinson, 1950). This makes a multilevel approach especially appropriate, since students are nested in classes, each linked to a science teacher, and classes are nested in schools.

At the same time, teaching practices are embedded within broader school and territorial contexts. In the specific case of Italy, macroarea differences provide an important analytical dimension. The macroarea variable distinguishes five territorial areas: North–West, North–East, Centre, South, and South and Islands. Educational opportunities and student outcomes may vary considerably across these areas. The present contribution focuses on the Italian Grade 4 case and examines whether ac–tion–based environmental instruction is associated with students' Environmental Knowledge after ac–counting for relevant student– and context–level characteristics. Discussion–based practices are includ–ed as a further dimension of reported instruction. Given the territorial structure of the Italian education system, the study also investigates whether the role of action–based practices varies across macroareas, connecting the analysis of teaching practices with an equity–oriented reading of primary education.

Data used

The study uses TIMSS 2023 Grade 4 data for the Italian sample. The international microdata were downloaded from the IEA data repository and merged using IEA IDB Analyzer. Following the TIMSS data structure, variables from the student, teacher, school, and home questionnaires were linked to the student–level file. The territorial variable Areageo_5, identifying the Italian macroarea, was added from the open dataset made available through the INVALSI Statistical Service data cata–logue. The resulting analytical file combines students' cognitive data with contextual information from the TIMSS questionnaires and the Italian territorial classification.

Methodology

The full Italian TIMSS 2023 Grade 4 sample included 4,453 students, 266 classes, and 152 schools. Since each sampled class was linked to one science teacher, the intermediate level is referred to as the class level, while acknowledging that teaching–practice indicators were reported by science teachers. The complete–case analytical sample included 3,960 students, nested in 255 classes and 147 schools. This sample retained students with valid information on Environmental Knowledge plausible values, sampling weights, student, class and school identifiers, gender, socioeconomic background, macroarea, and the teaching–practice indicators. The same analytical sample was used across the nest–ed models.

Teaching–practice indicators were constructed from the TIMSS 2023 science teacher questionnaire. Discussion–based environmental instruction (DBEI) was based on four reverse–coded frequency items asking teachers how often they address environmental topics in class. A weighted confirmatory factor analysis supported a one–factor solution (CFI = .995; RMSEA = .077; SRMR = .012; ω = .887), with standardized factor loadings ranging from .693 to .908. Action–based environmental instruction (ABEI) was based on four dichotomous items referring to visits to natural areas, environmentally re–sponsible activities, research projects on environmental topics, and participation in outdoor environ–mental education programs. A weighted CFA using WLSMV for binary indicators also supported a one–factor solution (CFI = .986; RMSEA = .033; SRMR = .047; ω = .65), with standardized loadings ranging from .607 to .814. CFA–based factor scores were used to construct class–level indicators. To distinguish class–level and school–level associations, each teaching–practice indicator was decomposed into within–school and between–school components. The within–school

component was computed as the class score minus the corresponding school mean, while the between–school component was com–puted as the school mean centered around the grand mean. School means were calculated on a class–level dataset, with one row per sampled class, so that they did not reflect differences in class size. This specification allowed the models to separate variation between classes within the same school from variation between schools.

Student– and context–level covariates included gender, socioeconomic background, and macroarea. Gender was recoded as a binary indicator, with male students as the reference category. Socioeco–nomic background was measured using the TIMSS socioeconomic scale and centered around the grand mean. Macroarea was included through the five–category Italian variable Areageo_5, with North–West as the reference category. Descriptive checks showed modest correlations among predic–tors (maximum |r| = .395), and VIF diagnostics did not indicate multicollinearity problems, with a maximum VIF of 1.09 in the main model and 1.92 in the interaction model.

The main analyses were conducted using three–level multilevel models, with students nested in classes and classes nested in schools. Environmental Knowledge was represented by the five TIMSS plausi–ble values. Models were estimated separately for each plausible value and then combined using Ru–bin's rules. Level–specific sampling weights were applied at the student, class, and school levels to support population–based estimates. The modelling strategy proceeded stepwise: an unconditional model, a model including macroarea, a model adding student–level covariates, and a main model in–cluding the decomposed teaching–practice indicators. Two exploratory extensions were then estimat–ed: an interaction model testing whether the within–school association between ABEI and Environ–mental Knowledge varied across macroareas, and a random–slope sensitivity model allowing this as–sociation to vary across schools. Analyses were conducted in R, using lavaan for confirmatory factor analyses and WeMix for weighted multilevel models. As this is a preliminary modelling stage, repli–cate–weight variance estimation will be incorporated in the final version of the study.

Results and Discussion

The multilevel analyses indicate that Environmental Knowledge varies across school contexts. In the unconditional model, the contextual ICC is .176, with most of this variance located at the school level. When macroarea is introduced, the contextual ICC decreases to .125, pointing to a substantial territo–rial component in between–school variation. Adding student–level covariates further reduces the con–textual ICC to .103. Socioeconomic background is positively associated with Environmental Knowledge ($B = 14.7$, $SE = 0.982$, $p < .001$), while girls score lower than boys ($B = −14.2$, $SE = 4.53$, $p = .012$). Although gender is not the central focus of the study, this result raises a further educa–tional question about whether environmental and science–related activities offer the same opportunities for participation and learning to all pupils. Compared with students in the North–West, those in the South ($B = −24.3$, $SE = 7.87$, $p = .006$) and in the South and Islands ($B = −45.5$, $SE = 10.4$, $p < .001$) obtain significantly lower scores even after controlling for socioeconomic background and gender. In the main model, the contextual ICC is further reduced to .102, with an approximate reduction of 17.9% in total variance and 51.9% in school–level variance relative to the unconditional model. These results place Environmental Knowledge within the broader territorial patterns observed in the Italian education system and indicate that environmental learning is also shaped by unequal educational con–ditions across territories and schools.

After separating teaching–practice indicators into within–school and between–school components, the key result concerns action–based environmental instruction. Its within–school component is positively associated with Environmental Knowledge ($B = 23.9$, $p < .001$), meaning that, within the same school, classes with higher teacher–reported levels of action–based instruction than the school average tend to obtain higher student scores. The between–school component is not significant ($B = 1.18$, $p = .885$), so the association appears to be located at the class level within schools, not at the level of school averages. The relevant difference, therefore, lies in the concrete learning experiences students encounter in class. Environmental Knowledge appears to be more closely linked to hands–

on envi–ronmental activities carried out at class level than to the school's overall average level of action–based instruction. This gives support to an action–oriented reading of environmental and sustainability edu–cation, where activities such as visits to natural areas, research projects, environmentally responsible initiatives, and outdoor programs appear to matter when they connect environmental topics with stu–dents' direct experience and scientific learning.

Discussion–based environmental instruction is not significantly associated with Environmental Knowledge in the main model, either at the within–school level ($B = -2.78$, $p = .624$) or at the be–tween–school level ($B = -6.73$, $p = .303$). This result leaves open the possibility that classroom dis–cussion is more closely related to Environmental Knowledge when it is connected to scientific con–tent, concrete environmental tasks, and opportunities for students to work actively with environmental issues. Discussion may also help students develop a more reflective relationship with environmental problems, while its connection with the factual and conceptual knowledge captured by TIMSS may depend on how it is structured in classroom practice. A plausible educational reading is that discussion becomes more relevant for knowledge when it is part of a wider learning sequence in which students observe, investigate, act, and reflect on environmental phenomena.

The interaction model did not provide clear evidence that the within–school association between ac–tion–based instruction and Environmental Knowledge varies systematically across macroareas, since all interaction terms were non–significant ($p \geq .194$). The random–slope sensitivity model, estimated on the subset of schools with sufficient within–school variation in action–based instruction, is consistent with the main result. The within–school component of action–based instruction remains positively as–sociated with Environmental Knowledge, although with greater uncertainty ($B = 25.1$, $SE = 12.4$, $p = .042$). The between–school component is not significant, and discussion–based instruction remains un–related to Environmental Knowledge. The estimated random–slope variance was non–negligible across plausible values, ranging from 5,036 to 7,431, and is consistent with possible between–school hetero–geneity in the strength of the association between action–based instruction and Environmental Knowledge. This leaves open the question of which classroom and school conditions allow hands–on environmental activities to become meaningful opportunities for learning. The random–slope result al–so suggests that the potential of action–based environmental instruction may not be equally available across schools. In educational terms, this variation brings attention to the school and classroom condi–tions under which these activities become regular science learning opportunities for students across classes.

Conclusions

These findings bring attention to how environmental and sustainability education becomes a learning opportunity in the classroom. The results call for closer attention to the ways in which environmental topics are made workable for pupils in everyday science teaching. Activities that bring pupils outside the classroom, involve local environmental investigation, or connect school work with everyday prac–tices should be recognized as part of the educational substance of ESE. They also point to the im–portance of teacher professional development in helping teachers build classroom activities in which direct experience and reflection are connected to scientific content. Territorial and socioeconomic dif–ferences make it difficult to assume that all schools are in the same position to offer these experiences. In some contexts, environmental activities may be harder to organize because project resources are limited, relationships with local actors are less stable, and outdoor spaces or materials are not easily available. Support for ESE should be sensitive to these differences and should also consider the pro–fessional and organizational conditions that sustain environmental learning at school level. The dis–tinction between class–level and school–level indicators also has methodological and policy relevance, since monitoring the presence of environmental projects at school level may hide important differ–ences in what pupils actually encounter across classes. For this reason, future research and system–level indicators should pay closer attention to how often classroom activities occur, whether they are sustained over time, and what kind of pedagogical work they involve. Sustained attention to these practices, especially in territorially and socially more

fragile contexts, may help sustainability educa–tion respond more concretely to equity concerns and to the ambitions set out in international frame–works.

Keywords: Environmental and Sustainability Education; TIMSS 2023; Environmental Knowledge; Action–Oriented Education

Purifying Calibration Samples under Aberrant Responding: A Likelihood–Ratio Approach to Survey Responses in ICCS

David Torres Irribarra – Camila Echeverría – Pablo Espinoza – Diego Carrasco

Introduction

Item calibration in large–scale assessment rests on the assumption that responses reflect the intended response process. Two well–documented departures threaten this assumption: random or quasi–random guessing on difficult multiple–choice items in high–stakes cognitive tests, and careless or insufficient–effort responding on questionnaire scales administered under low–stakes conditions — the setting characteristic of international large–scale assessments (ILSAs). Both are content–independent response processes that, left unmodelled, bias item parameter estimates, distort item characteristic curves, and compromise the linking and cross–system comparability on which policy uses of ILSA results depend.

The methodological literature offers increasingly rich models for these processes in academic tests: lower–asymptote parameters, ability–based guessing models (San Martín, del Pino, & De Boeck, 2006), mixture models (Mislevy & Verhelst, 1990), response–time and engagement models (Wise & DeMars, 2005), and response–level elimination procedures (Andrich, Marais, & Humphry, 2016). For questionnaire data, a parallel literature on careless and insufficient–effort responding has developed post–hoc indices such as longstring, intra–individual response variability, and Mahalanobis distance (Meade & Craig, 2012; Curran, 2016; Dunn, Heggestad, Shanock, & Theilgard, 2018). Richer models, however, are frequently infeasible within operational calibration timelines, and index–based screening lacks an explicit model of the aberrant process being removed.

This paper presents a deliberately simple alternative — person–level purification of the calibration sample — and applies it to survey responses in the International Civic and Citizenship Education Study (ICCS), with citizenship competencies as a substantively important and methodologically underexplored non–cognitive outcome. The extension builds directly on our previous work with ICCS data. Using ICCS 2016, we identified cross–national profiles of citizenship norm endorsement via multigroup latent class analysis (Treviño, Carrasco, Claes, & Kennedy, 2021), and subsequently examined whether the extreme profiles in that solution reflect substantive positions or response artifacts, using confirmatory latent class models with constrained artifact classes (uniform endorsement, uniform rejection) evaluated against careless–responding indices (Torres Irribarra, Carrasco, Ponce, & Escribano, 2023). That work established the latent–class machinery for separating substantive from artifactual response patterns in ICCS questionnaire data; the present contribution develops its operational counterpart, in which classification serves calibration rather than profiling. The contribution is not another model of disengagement, but an operational procedure for obtaining cleaner calibration samples under realistic production constraints.

Data used

Two data sources are used. First, a 48–condition Monte Carlo simulation (60 items, 10,000 examinees, 100 replications per condition) crossing the prevalence of aberrant responders (5%,

15%), the proportion of affected items (20%, 50%, 80%), the heterogeneity of item guessing floors, and the degree to which aberrance concentrates at low ability, with parameter recovery scored against the guessing–free generating values. Second, ICCS student questionnaire and civic knowledge data (ICCS 2016 for continuity with our prior latent–class work, and ICCS 2022 for the planned analyses) with Chile and Italy as focal systems for cross–country comparison, responding to the session's interest in reflections on Italian data relative to other educational systems.

Methodology

The purification procedure compares, for each respondent, the likelihood of the observed response vector under the operational measurement model against its likelihood under an explicit aberrant–response model, and removes respondents classified as aberrant from the calibration sample only; all respondents are subsequently scored under the purified item parameters. The procedure amounts to a likelihood–ratio classifier followed by recalibration — the hard–assignment limit of a two–class mixture model (Mislevy & Verhelst, 1990), which converges with soft assignment when class separation is high (Celeux & Govaert, 1992).

This ICCS application replaces the guessing floor with generating classes appropriate for Likert–type responses, namely uniform category endorsement, straightlining, and extreme–response patterns. These are evaluated against the partial credit trait models used in the operational scaling of ICCS questionnaire scales, alongside a direct application of the original procedure to the ICCS civic knowledge test. These generating classes correspond to the constrained artifact classes of our confirmatory latent class work (Torres Irribarra et al., 2023).

Evidence supporting the classification relies on the recurrence of aberrant patterns across questionnaire sections. Classification rates, consequences for scale calibration, and the behaviour of the procedure are compared between Chile and Italy.

Results and discussion

The analyses are in progress, and we anticipate three classes of findings. First, classification rates for artifact–consistent responding on civic attitude scales comparable to prevalence estimates in the careless–responding literature (Meade & Craig, 2012) and to the constrained–class prevalences observed in our previous ICCS analyses. Second, heterogeneity in these rates across countries, with Chile and Italy as the focal contrast; to the extent rates differ, we expect calibration consequences concentrated at the extremes of the scale locations, where cross–country comparability claims are most exposed. Third, and most importantly, the analyses will establish whether disengaged responding in ILSA questionnaires exhibits the concentrated structure the method requires.

Conclusions

Differential disengagement across countries is an important concern for the comparability claims on which policy uses of ILSAs rest. If the anticipated findings hold, a fast, transparent, auditable purification of the calibration sample — requiring only the response matrix — would offer a practical complement to invariance testing and to heavier process–data approaches, with boundary conditions mapped in advance rather than discovered in application. The anticipated contribution is twofold: an empirical test of whether a procedure with simulation–established properties transports to low–stakes, non–cognitive measurement in ICCS, and evidence on the structure of disengaged responding in citizenship competency scales across educational systems. We will discuss implications for the calibration and cross–system comparison of these scales, and the policy conditions under which information about likely disengaged respondents can be used responsibly.

Keywords: International large–scale assessments; Careless responding; Item calibration; ICCS

Using International Large–scale Assessments to Measure Action Competence and Action Orientation

Adam Coates

Introduction

The growing urgency of climate change has prompted widespread initiatives to combat environmental issues. At the start of the 21st century, a growing consensus emerged that lasting solutions to climate change require cultural and behavioural changes rather than just economic adjustments (WHO, 2011). This shifted the spotlight to education as a primary instrument for change, triggering a global expansion of sustainability learning policies and frameworks (Hadjichambis et al., 2020).

While global adoption of sustainability education has progressed, systematically monitoring and evaluating its implementation remains a major challenge (UNESCO, 2019). Current oversight by bodies like UNESCO and the Eurydice network focuses heavily on intended inputs, such as curriculum content and teacher training guidelines (European Commission / EACEA / Eurydice, 2024). Consequently, the monitoring of actual classroom processes and student outcomes remains underdeveloped, failing to capture how these policies are enacted or how they vary within school systems (UNECE, 2025).

Although not explicitly designed for tracking sustainability education processes and outcomes, data from international large–scale assessments offer great potential for this purpose. For instance, UNESCO and the European Commission leverage findings from the International Civic and Citizenship Education Study (ICCS) to evaluate student knowledge, values, and environmental actions (European Commission: Directorate–General for Education, Youth, Sport and Culture, 2024; UNESCO, 2025). Additionally, researchers have used data from the Trends in International Mathematics and Science Study to measure students' overall environmental knowledge (Bookbinder et al., 2025; Isac et al., 2025).

Action orientation and action competence

Most sustainability education curricula focus on building knowledge and behaviours (UNESCO, 2019). However, recent frameworks like *GreenComp* shift this focus, aiming to equip students with the willingness and capacity for self–directed action across individual, collective, and political spheres (Bianchi et al., 2022). By integrating these distinct lenses, modern approaches explicitly elevate action competence as a foundation of environmental learning.

Action competence is defined as an individual's current and future capacity to take responsibility and engage in self–directed action rather than blindly adopting prescribed behaviours (Jensen & Schnack, 1997). This concept has been integrated into sustainability education to empower learners to actively address environmental issues. Recently, Sass et al. (2020) redefined the concept, arguing that true competence requires a combination of capability, motivation, knowledge, and confidence. To capture this, they proposed a framework divided into three distinct dimensions:

Knowledge of action possibilities Knowledge of action possibilities combines the understanding of potential solutions, stakeholders, and affected groups with essential skills like critical, strategic, and interpersonal thinking.

Confidence in one's own influence Confidence in one's own influence merges self–efficacy—the personal belief that one can create change—with outcome expectancy, which is the expectation that these personal efforts will achieve the intended results.

Willingness to act Willingness to act balances an individual's intrinsic motivation and identity–driven desire to act sustainably against their readiness to accept the associated economic, social, or personal effort costs.

Because fostering action competence departs from traditional educational goals, teachers must adopt specialized pedagogical approaches. To guide this shift, a recent literature review of sustainability education theory and research identified five core principles for action–oriented teaching and learning (Karaarslan Semiz et al., under embargo). These fundamental principles provide a direct framework for designing instruction that actively builds action competence in students. The five pedagogical principles are learning–through–action, real–world integration, democratic participation, critical reflection, and emotional awareness.

This study tests the possibility of creating indicators for dimensions in the action competence and action orientation frameworks based on data from international large–scale assessments. Validating the use of international large–scale assessments to measure the processes and outcomes of sustainability and climate change education will provide a cost–effective strategy to repurpose existing data for monitoring sustainability and climate change learning.

This approach addresses a critical gap by shifting the evaluative focus away from inputs, such as curriculum guidelines, and toward tangible classroom processes and student outcomes. As a result, the standardized indicators can provide a robust empirical baseline to track the efficacy of their sustainability programs across education systems. These indicators will allow countries to identify relative weaknesses so that strategic interventions can be designed and resources allocated to efficiently strengthen the targeted areas of their school systems.

Data used

The results in this presentation use data from the 2022 International Civic and Citizenship Education Study (ICCS) for European countries. Items from student questionnaire IS4G15, IS4G16, IS4G19, IS4G29, and IS4G31 and from teacher questionnaire IT4G11 and IT4G14 were used to create scales.

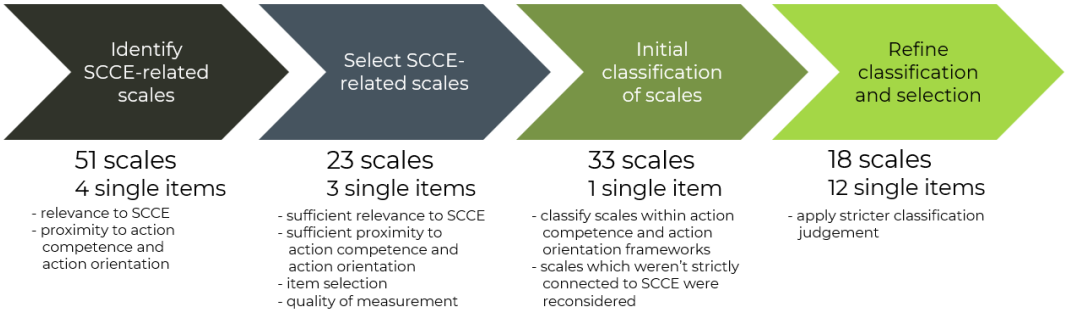
Methodology

This study presents the results of linking International large–scale assessment with action competence and pedagogical principles for action orientation.

Figure 1 outlines the process through which potential scales were selected. Through this process, 18 scales and 12 single items were ultimately chosen for further analysis.

Figure 1

Process of identifying and classifying ILSA scales



Note: a scale refers to two or more questions in a battery of questions

The psychometric robustness of the selected scales was then tested. This involved evaluation of internal consistency; item–whole correlations; dimensionality using both parallel and confirmatory factor analysis; invariance testing; and IRT model fit statistics.

For the scales that failed to show strict or scalar invariance, alignment optimisation (Asparouhov & Muthén, 2014) was used to identify parameters which showed significant non–invariance, defined as being non–invariant for more than 25% of groups. These parameters were then freed in a partial invariance model, with a cap of allowing a maximum of 25% of parameters to be free.

Results and discussion

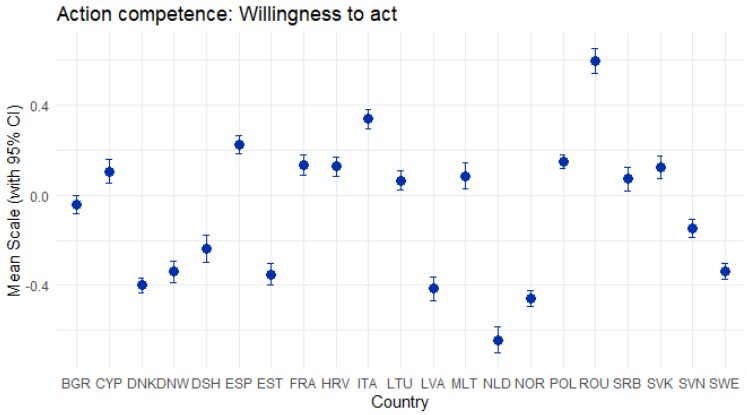
This presentation focuses on the five student scales using ICCS data which were sufficiently robust and allowed comparisons across countries. Details of these scales are displayed in Table 1.

Table 1
Model fit of the scales

Scale	ICCS scale	Model	CFI	TFI	RMSEA	SRMR
Willingness to act	IS4G31	Scalar	0.981	0.990	0.083	0.031
Confidence in one's own influence	IS4G29	Partial Scalar	0.978	0.988	0.045	0.033
Democratic participation	IS4G15	Partial Scalar	0.967	0.988	0.046	0.030
Democratic participation	IS4G16	Partial Scalar	0.976	0.988	0.035	0.037
Democratic participation	IS4G19	Partial Scalar	0.967	0.988	0.046	0.030

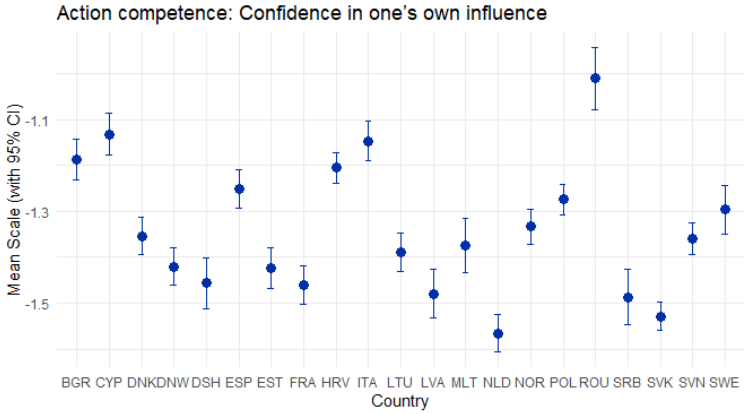
Figures 2 and 3 show the country mean student scores on the two action competence scales: *willingness to act* and *confidence in one's own influence* scales. There is some degree of consistency among the highest and lowest scoring countries across the scales: Romania and Italy show high average scores on both scales, while the Netherlands, Latvia, and France have the lowest mean scores. The correlation between the two scales was 0.32 on average, with some variation across countries.

Figure 2
Country averages on the willingness to act scale



Note: error bars represent the 95% confidence interval of the mean

Figure 3
Country averages on the confidence in one's own influence scale



Note: error bars represent the 95% confidence interval of the mean

Data allowed for the construction of scales for only one of the identified pedagogical principles. The three democratic participation scales measure students' perceptions of various school activities: IS4G15 relates primarily to formal student representation in decision–making; IS4G16 relates to an

open classroom climate where students can freely discuss ideas; and IS4G19 relates to student representation within the school system. Although these scales were anticipated to measure the same construct, correlations were below 0.5 at both individual and country–mean levels. This indicates a theoretical issue requiring further analysis.

Importantly, democratic participation is expected to support the development of action competence. Table 2 presents the regression coefficients for linear regressions of the action competence scales predicted by the three standardized democratic participation scales. The results suggest that the scale based on IS4G15 is the most viable choice, consistent with the theoretical assumption that a significant relationship should exist between these variables.

Table 2

Correlations between democratic participation and action competence scales

	IS4G15	IS4G16	IS4G19
Willingness (IS4G31)	0.166	0.154	0.094
Confidence (IS4G29)	0.286	0.064	0.088

Conclusion

The results of this study demonstrate the potential for using ILSA data to monitor action competence and action orientation. Future steps involve analysing earlier cycles of ICCS, further exploring the relationships between different aspects of the frameworks, and combining the results from ICCS with those from other ILSAs.

Using a tool for a goal other than its intended purpose will always bring challenges. Trying to fit ILSA data into the action competence and orientation frameworks results in two main obstacles. First, not all concepts in the theoretical frameworks can be captured in ILSA data. In the ICCS student data, for example, background questionnaire items aligned with only one action orientation principle. Second, even when items do match the theoretical frameworks, the alignment can be weak. While the willingness to act measure is specifically tied to climate change and sustainability actions, the measures of willingness to act and democratic participation are more general – associated more broadly with citizenship. This, in part, may explain why the relationships between the scales may not strongly match the theoretical expectation.

Keywords: Action competence; Action orientation; Sustainability and climate change education; ICCS

Vertical and Horizontal Inequalities in Students' Educational and Occupational Expectations: A Comparative Analysis of HRSTO Indicator in Spain, Italy, and Portugal based on PISA 2022 Data

Juliana Hartz Rodrigues – Enrico Nerli Ballati

Introduction

Vertical inequalities refer to differences in the level of education or social position that individuals are expected to reach, while horizontal inequalities refer to differences in the type, field, or qualitative orientation of trajectories among individuals who may appear similar in vertical terms. In education, vertical inequality can be observed when students from different social backgrounds expect to reach different levels of schooling. Horizontal inequality, by contrast, may emerge when students with similarly high educational expectations imagine different kinds of occupational futures, some of which are more socially valued, prestigious, or strategically positioned in the labour market.

This distinction is particularly relevant for the Spanish case, where previous research has shown that educational expansion may reduce some forms of vertical inequality while leaving room for new forms of horizontal differentiation in educational trajectories (Valdés, 2022).

The study examines how social origin is associated with students' expected educational and occupational trajectories, using public–use microdata from PISA 2022. The initial focus is Spain, but the study also proposes a comparative extension including Italy and Portugal. This comparative design allows the analysis to explore whether the relationship between socioeconomic background, educational expectations, and occupational expectations varies across closely related Southern European contexts.

Spain, Portugal, and Italy share a common background of late educational expansion and historically high levels of early leaving from education and training. At the same time, the three countries differ substantially in tertiary educational attainment among 25–34 year–olds. Spain stands clearly above the EU average, with 52.6% of young adults having completed tertiary education in 2024, while Portugal has almost converged with the EU average, reaching 43.2% compared with 44.1% in the EU. Italy, by contrast, remains behind, with only 31.6% of 25–34 year–olds holding tertiary education (Eurostat, 2025b).

These similarities and divergences make the comparison analytically useful: the three countries share a Southern European background, but they occupy different positions in the expansion of tertiary education and in the development of educational and occupational pathways linked to science, technology, and HRST–related occupations (Eurostat, n.d.). Analyzing Spain, Portugal, and Italy together makes it possible to examine whether potential vertical and horizontal inequalities follow a shared Southern European pattern, while also identifying national differences in educational expectations and HRST–related occupational aspirations.

The study focuses on potential rather than observed inequality. PISA does not follow students into adulthood and therefore cannot measure their actual educational or occupational outcomes. However, it provides valuable information about students' expectations at age 15, a crucial point at which future trajectories are already being imagined, evaluated, and socially structured. The central argument is that inequalities may emerge before actual transitions take place, through students' expectations about education and work, following the proficiency levels, effects of students' social origin as well as structural contexts.

Data used

The analysis uses the PISA 2022 public–use microdata, specifically the student questionnaire data file. This file is appropriate for the study because it contains students' educational expectations (`EXPECEDU`), occupational expectations (`OCOD3`), socioeconomic background (`ESCS`), student weights, replicate weights, and plausible values for achievement (OECD, 2023a).

The classification of occupational expectations follows Eurostat's HRST framework: HRSTO refers to science and technology occupations, while the more restrictive SE category identifies scientists and engineers (Eurostat, n.d.).

As previously mentioned, the analysis focuses on Spain, Portugal, and Italy. In PISA 2022, Spain assessed 30,800 students in 966 schools, representing about 459,000 15–year–old students, or around 90% of the total population of 15–year–olds. Portugal assessed 6,793 students in 224 schools, representing about 96,600 15–year–old students, or around 93% of the total population. Italy assessed 10,552 students in 345 schools, representing about 496,300 15–year–old students, or around 87% of the total population (OECD, 2023b, 2023c, 2023d).

PISA 2022 is designed to represent 15–year–old students enrolled in educational institutions at grade 7 or higher, with national samples representing the target population in each participating country (OECD, 2023e).

The analysis also draws on national PISA documentation produced by INEE for Spain, IAVE for Portugal, and INVALSI for Italy, which provides country–specific information on implementation,

reporting, and sample design (IAVE, 2023; INVALSI, 2023; Ministerio de Educación, Formación Profesional y Deportes, 2023).

In all three cases, PISA provides nationally representative samples of 15–year–old students enrolled in school. Spain, however, uses a substantially larger sample than Portugal and Italy because the Spanish design includes an enlarged sample that is representative not only at the national level but also across its autonomous communities and autonomous cities.

The PISA–based estimates of HRSTO–related occupational expectations among 15–year–olds will also be compared with the official Eurostat HRSTO indicator for the same three countries in 2022. While PISA captures expected future occupations among students, Eurostat's HRST indicators describe the adult labour force and are derived from the European Union Labour Force Survey. This comparison is not intended to equate expectations with realised labour–market outcomes, but rather to assess how closely students' expected occupational futures resemble the actual structure of science–and–technology–related employment in Spain, Portugal, and Italy in 2022 (Eurostat, 2026a, 2026b).

Methodology

Potential vertical inequality is measured through the PISA variable EXPECEDU, which captures the highest level of education students expect to complete. This variable is recoded according ISCED classification into indicators distinguishing students who expect to reach tertiary education, those who expect to reach university–level education or above, and those who do not expect tertiary education. Social origin is measured primarily through the PISA index of economic, social and cultural status, ESCS, a composite indicator that is widely adopted and plays a strategic role in the PISA framework, as it summarises multiple dimensions of family background. Potential horizontal inequality is measured through students' expected occupations. The variable OCOD3, which records students' expected occupation coded according to ISCO categories, is used to classify occupational expectations into Human Resources in Science and Technology Occupations, or HRSTO.

Following Eurostat's HRST framework, HRSTO refers to individuals employed in science and technology occupations, identified through ISCO occupational classifications and measured in official statistics using Labour Force Survey data. In the PISA–based analysis, students are classified as having a potential HRSTO expectation when their expected occupation belongs to ISCO major groups 2 or 3, corresponding broadly to professionals, technicians, and associate professionals. A more restrictive indicator is also created for potential scientists and engineers, identifying occupations in ISCO groups 21, 22, and 25. Special response codes in OCOD3, such as no response, "I do not know", vague answers, and invalid codes, are treated separately rather than being classified as non–HRST occupations.

The main research hypothesis is that students from higher socioeconomic backgrounds will display more ambitious expected trajectories than students from lower socioeconomic backgrounds. This is expected to appear both vertically, through higher probabilities of expecting tertiary or university–level education, and horizontally, through a greater likelihood of anticipating HRSTO or scientist–and–engineer occupations. The study further expects horizontal inequality to remain visible even among students who already expect to reach tertiary education, following the argument that qualitative differentiation may persist even when vertical educational access expands (Valdés, 2022).

Expected results and discussion

A key contribution of the study is that it produces an estimate of HRSTO–related occupational expectations among 15–year–olds using PISA 2022 data, while also comparing this estimate with the official Eurostat HRSTO indicator for Spain, Portugal, and Italy in 2022, derived from the Labour Force Survey. This comparison does not equate students' expectations with realized labour–market outcomes. Rather, it allows the analysis to examine whether young people's expected occupational destinations resemble or diverge from the actual structure of science–and–technology–related employment in each country.

Country differences are also expected in the balance between vertical and horizontal inequalities. In Spain, relatively high tertiary attainment suggests that vertical educational expectations may be broadly widespread, but horizontal differentiation may be especially important if students from advantaged backgrounds are more likely to expect HRSTO or scientist–and–engineer occupations. In Portugal, where tertiary expansion has progressed substantially and early leaving has declined sharply, both vertical and horizontal inequalities may appear in a transitional form: expectations of higher education may have become more generalized, while access to more selective occupational imaginaries may still vary by social origin. In Italy, where tertiary attainment remains comparatively low, vertical inequality may be more pronounced, with social origin strongly shaping whether students expect to reach higher education at all; however, horizontal inequality may also emerge among those who do expect tertiary education, particularly in relation to HRSTO and science–and–technology–related occupational aspirations.

Conclusions

This study brings together subjective expectations and official population–level indicators within a common comparative framework. By using PISA 2022, the analysis captures how 15–year–old students imagine their future educational and occupational trajectories before actual transitions into higher education and the labour market occur.

By comparing these expectations with Eurostat HRSTO estimates derived from the 2022 Labour Force Survey, the study connects young people's anticipated futures with the observed structure of science–and–technology–related employment in each country.

The study also benefits from international comparability: PISA provides microdata on students' expectations and socioeconomic background, while Eurostat provides official indicators of tertiary attainment and HRST across countries. A further strength is the application of official statistical classifications, particularly ISCO–based HRSTO and scientist–and–engineer categories, to high–quality survey data on students' expected occupations. In doing so, the study offers an original contribution to research on educational inequality by analysing not only whether students expect to reach higher education, but also whether they imagine access to strategically valued occupational destinations in science and technology.

Notes

This study is developed within the framework of the Istat curricular internship project "027 – Surveys on the Education System", under the area "Methods and Techniques for Statistical Surveys." The study aligns closely with the aims of the internship by combining high–quality international microdata, official statistical classifications, and comparative indicators to investigate educational and occupational expectations.

Keywords: Vertical and horizontal inequalities; Occupational expectations; PISA 2022; HRSTO indicator

Organisational Conditions and Teachers' Use of Digital Resources and IA in European Schools: The Mediating Role of Teacher Collaboration, Professional Learning, and Self–Efficacy: A Cross–National Analysis Based on TALIS 2024

Angelo Paletta – Valentina Liguori – Daniele Vidoni

The integration of digital tools and artificial intelligence (IA) in teaching practice has become a central policy priority across European school systems (European Commission, 2020; OECD, 2025), yet adoption rates among teachers remain highly uneven within and across countries (Tondeur et al., 2017). Drawing on Professional Learning Community (PLC) theory and organisational learning theory, this study examines the following research question:

In what ways do school–level organisational conditions – specifically collaborative governance and distributed leadership – shape teacher collaboration and teachers' professional learning in artificial intelligence and digital tools, and how do these in turn predict teachers' self–efficacy and actual use of digital resources in teaching across European school systems?
The study draws on data from the Teaching and Learning International Survey (TALIS) 2024 and employs a Multilevel Structural Equation Modelling approach to test a parallel and sequential mediation model. Findings are expected to demonstrate that the pathway from organisational conditions to AI adoption operates through two parallel mechanisms – teacher collaboration and structured professional learning in AI and digital tools – which jointly strengthen teacher self–efficacy and, in turn, predict actual digital use in teaching.

Purpose

The integration of digital tools and artificial intelligence (IA) in teaching practice has attracted increasing policy attention across European educational systems, driven in part by the European Digital Education Action Plan[1] (2021–2027) and by growing recognition of digital innovation and AI as a transformative forces in public service delivery (European Commission, 2020; OECD, 2025). Despite sustained investment in digital infrastructure and teacher training, however, adoption rates remain highly uneven across and within countries, and the conditions that determine whether teachers effectively integrate digital tools and AI into their instructional practice are not yet well understood (Tondeur et al., 2017; Voogt et al., 2018). Around one in three teachers report having used AI in their work across OECD education systems, with rates ranging from 75% in Singapore and the United Arab Emirates to fewer than 20% in France and Japan (OECD, 2025).
Existing research has predominantly examined individual–level predictors of AI adoption – including teacher age and prior experience (Voogt et al., 2018), digital self–efficacy (Ertmer & Ottenbreit–Leftwich, 2010), and beliefs about the benefits and risks of AI in teaching (Tondeur et al., 2017) – while systematically neglecting the organisational conditions that shape these individual orientations (Ertmer & Ottenbreit–Leftwich, 2010; Tondeur et al., 2017).
Drawing on Professional Learning Community (PLC) theory (Hord, 1997; Stoll et al., 2006) and organisational learning theory (Senge, 1990; Leithwood & Louis, 2012), this study proposes that school–level collaborative governance and distributed leadership function as organisational enabling conditions that foster teacher collaboration and create structured opportunities for professional learning in AI, which jointly strengthen teacher self–efficacy and predict actual AI use in teaching.
The study addresses the following research question: *In what ways do school–level organisational conditions – specifically collaborative governance and distributed leadership – shape teacher collaboration and teachers' professional learning in artificial intelligence and digital tools, and how do these in turn predict teachers' self–efficacy and actual use of AI in teaching across European school systems?*
Based on the theoretical framework outlined above, the following hypothesis are proposed (see Figure 1):
H1: Collaborative governance is positively associated with teacher collaboration and professional learning in AI and digital tools.
H2: Distributed leadership is positively associated with teacher collaboration and professional learning in AI and digital tools.
H3: Teacher collaboration is positively associated with teacher self–efficacy in digital tools.

[1] The Digital Education Action Plan (2021–2027) is the European Commission's overarching policy framework for supporting the adaptation of education and training systems to the digital age across EU member and associated states. It sets two strategic priorities: fostering the development of a high-performing digital education ecosystem, and enhancing digital skills and competencies for the digital transformation. Full text available at: https://education.ec.europa.eu/focus-topics/digital-education/action-plan

H4: Professional learning in AI and digital tools is positively associated with teacher self–efficacy in digital tools.
H5: Teacher self–efficacy in digital tools is positively associated with actual use of AI in teaching.

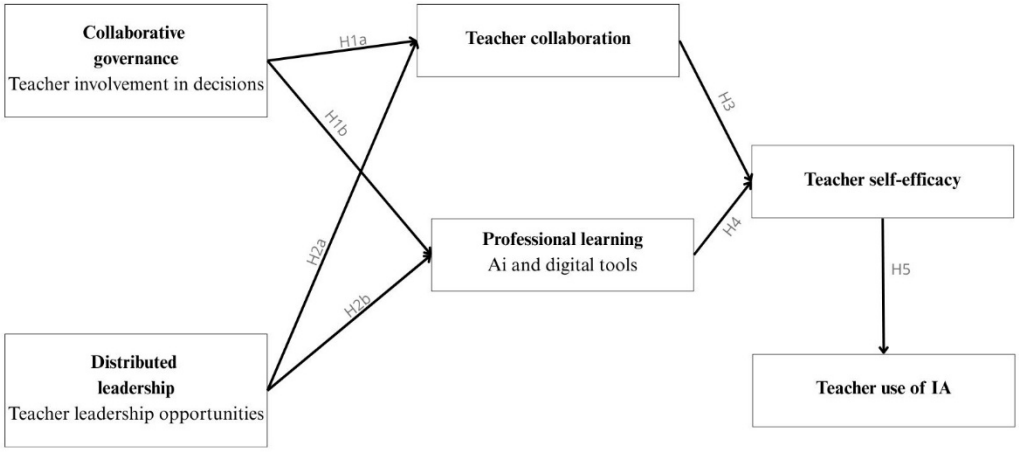


Figure1: Proposed research model.

Methodology

The study draws on data from the Teaching and Learning International Survey (TALIS) 2024, the largest international survey specifically designed to examine the professional conditions of teachers and school principals (OECD, 2025), covering 49 countries and territories across multiple school levels. The analysis focuses on lower secondary education (ISCED level 2), the level for which TALIS 2024 provides its broadest cross–national coverage (OECD, 2025). The analytical sample is further restricted to the European subsample, in order to ensure institutional comparability and policy coherence across the sample, given the common overarching framework provided by the European Digital Education Action Plan (European Commission, 2020).

The analytical model operationalises five categories of constructs derived from validated TALIS 2024 scales. At the school level, *collaborative governance* is operationalised as the degree of teacher involvement in school decision–making on school policies, and *distributed leadership* as teachers' perceived opportunities for leadership within the school – both constructs with established validity in the educational management literature (Spillane, 2006; Leithwood & Louis, 2012). At the teacher level, three mediating constructs are included: *teacher collaboration*, capturing the frequency and quality of collaborative professional practice; *professional learning in AI and digital tools*, capturing teachers' participation in professional development activities related to digital skills; and *self–efficacy in using digital tools and resources*, capturing teachers' confidence in integrating digital resources into instruction. The primary outcome variable is *teacher use of AI in teaching*, capturing the frequency of AI use in classroom practice.

A Multilevel Structural Equation Modelling (ML–SEM) approach is adopted to test the proposed model, accounting for the nested structure of teachers within schools and schools within national systems. The model estimates direct effects, sequential mediation pathways, and cross–level interactions between school–level organisational conditions and teacher–level outcomes. Sampling weights and variance estimation procedures consistent with the complex stratified cluster sample design of TALIS 2024 are applied throughout.

Results and Discussion

The present study reports preliminary findings based on bivariate correlation analysis of the TALIS 2024 European subsample. Full ML–SEM analysis, including the professional learning construct, is currently underway. Preliminary evidence provides initial support for the proposed organisational pathway. Collaborative governance and distributed leadership are positively associated with professional collaboration in lessons among teachers ($r = 0.18$ and $r = 0.24$, respectively), consistent with H1 and H2. Teacher collaboration, in turn, shows positive associations with self–efficacy in digital tools. The strongest association in the matrix is observed between teacher self–efficacy and actual use of digital resources in teaching ($r = 0.46$), providing robust initial support for H5. These findings suggest that the proposed organisational pathway is empirically plausible across the European educational context. It should be noted that these correlations represent preliminary estimates; full analysis incorporating TALIS 2024 BRR replicate weights and ML–SEM modelling is currently underway and may yield refined estimates.

Research Limitations

The cross–sectional nature of TALIS 2024 data precludes causal inference, and all measures rely on self–reported teacher and principal perceptions rather than direct observation of classroom practice. The restriction of the sample to the European subsample, while theoretically justified, limits the generalisability of findings to non–European school systems. Future research should employ longitudinal designs to establish causal pathways, and integrate student–level outcome data to link AI adoption to learning results.

Conclusions

Findings are expected to indicate that policies aimed at increasing AI adoption in schools cannot rely solely on individual teacher training and skill development. Organisational investment in collaborative governance structures, distributed leadership practices, and structured opportunities for teacher collaboration and professional learning communities is essential to create the conditions under which teacher self–efficacy and AI use develop sustainably. For school leaders, findings point to the importance of involving teachers in school decision–making and creating structured peer learning opportunities around AI.

This study contributes to the educational management and public management literature by advancing PLC theory through the identification of teacher collaboration, professional learning, and self–efficacy as the organisational mechanisms through which school–level conditions translate into instructional practice with digital tools. It provides the first cross–national European multilevel evidence on the organisational conditions that systematically predict digital resource use in lower–secondary school systems, leveraging the new digital resource variables introduced in TALIS 2024.

Keywords: Digital tools in education; Professional Learning Communities; Distributed leadership; Teacher self–efficacy; Teacher collaboration; TALIS 2024; Multilevel analysis; Organisational enabling conditions

SESSION 4. MEASURING AND DEVELOPING SOCIAL–EMOTIONAL COMPETENCIES IN SCHOOLS: METHODS, EVIDENCE, AND POLICY IMPLICATIONS

ORGANIZERS: POLITECNICO DI MILANO – UNIVERSITY OF NAPLES "PARTHENOPE"

COORDINATORS: GABRIELE IANNOTTA – SERGIO LONGOBARDI

NOVEMBER 4TH : 2.00 P.M. – 4.00 P.M. {AULA B0.1 – RESEARCH 3}

When Perseverance Overcomes Socio–Economic and Cultural Disadvantage: Predictors of Academic Resilience in Italy Using PISA 2022 Data

Paolo Barabanti

Introduction

Socio–economic background remains one of the strongest predictors of educational achievement across countries, raising persistent concerns about the capacity of education systems to ensure genuine equality of opportunity (Coleman et al., 1966; Sirin, 2005). Yet educational trajectories are not entirely determined by socio–economic disadvantage. A substantial minority of disadvantaged students consistently attain outstanding levels of performance despite adverse circumstances. Following the OECD framework, these students are defined as academically resilient: students whose socio–economic status falls within the lowest quartile of the national ESCS distribution while simultaneously performing in the highest quartile of academic achievement (OECD, 2011). As such, academic resilience has become one of the most widely adopted indicators for assessing whether education systems are capable of combining educational excellence with equity, identifying students who overcome structural disadvantage and successfully translate potential into high academic performance.

The growing attention devoted to academically resilient students reflects a broader shift in educational research. Rather than concentrating exclusively on the determinants of low achievement, scholars have increasingly investigated why some disadvantaged students follow unexpectedly successful educational trajectories and which individual, family, school, and institutional conditions enable these outcomes (Agasisti & Longobardi, 2014; Rudd et al., 2021). Accordingly, resilience is no longer conceptualised as a fixed individual trait but as a dynamic and multidimensional process emerging from the interaction between personal resources and the educational environments in which students develop (Luthar & Cicchetti, 2000; Masten, 2001; Rutter, 2012).

Current evidence portrays academic resilience as the product of multiple and interconnected protective factors operating across different ecological levels. At the individual level, attributes such as self–efficacy, motivation, perseverance, confidence, self–regulation and low anxiety have consistently been associated with resilient educational trajectories (Martin & Marsh, 2006). Beyond individual characteristics, contextual conditions also play a decisive role. Teacher and family support, positive school climate, cognitively activating instruction, orderly disciplinary environments and adequate educational resources all contribute to increasing the likelihood that socio–economically disadvantaged students achieve high levels of performance (Siebecke, 2024). Comparative studies further indicate that resilient students are disproportionately concentrated in schools characterised by favourable learning environments, lower levels of truancy, qualified teaching staff and stronger

organisational capacity, suggesting that resilience should also be interpreted as an indicator of how effectively education systems transform disadvantage into educational opportunity (Agasisti et al., 2021).

Despite the growing body of international research on academic resilience, evidence on the Italian education system remains relatively limited, particularly with respect to the combined role of individual and contextual factors in fostering resilient educational trajectories. Building on the OECD conceptualisation of academic resilience, this study aims to investigate the proportion and determinants of academically resilient students in Italy, situating the national case within the broader European context. More specifically, the study addresses the following research questions:

- RQ1. What proportion of students are academically resilient in Italy compared with the other European Union countries participating in PISA 2022?
- RQ2. Which socio–emotional and demographic characteristics distinguish academically resilient students from other socio–economically disadvantaged students in Italy?

Data used

The study uses data from the OECD Programme for International Student Assessment (PISA) 2022. The comparative analyses include all European Union countries participating in PISA 2022, whereas the multivariate analyses focus on the Italian student sample. Variables include students' achievement, socio–economic status (ESCS), gender, migration background, and socio–emotional skills (e.g., perseverance, emotional control, curiosity, empathy, cooperation, assertiveness, stress resistance, and sense of belonging).

Methodology

Academic resilience is operationalised according to the OECD definition, identifying students whose socio–economic status belongs to the lowest national quartile of the ESCS index while their mathematics achievement falls within the highest national quartile of the national performance distribution.

First, the proportion of academically resilient students is estimated across all European Union countries, allowing Italy to be positioned within the broader European context.

Second, the Italian sample is analysed to describe the socio–demographic profile of academically resilient students.

Finally, separate binary logistic regression models are estimated for Reading, Mathematics and Science. The analyses assess the contribution of socio–emotional skills, demographic characteristics and socio–economic status to the probability of belonging to the highest achievement quartile within each socio–economic group.

Results and discussion

The comparative analyses reveal substantial variation in the proportion of academically resilient students across European Union countries, highlighting important differences in the capacity of education systems to foster high achievement under conditions of socio–economic and cultural disadvantage.

Focusing on the Italian case, the regression analyses identify perseverance as the only socio–emotional skill consistently associated with academic resilience across all three cognitive domains. Higher levels of perseverance are associated with a greater probability of academic resilience across Reading, Mathematics and Science. Emotional control emerges as an additional predictor only in Science, whereas second–generation immigrant status is negatively associated with academic resilience in Mathematics and Science.

Overall, the findings suggest that perseverance represents the key socio–emotional resource associated with academic resilience, whereas the contribution of the other socio–emotional skills appears to be discipline–specific or statistically negligible.

Conclusions

This study contributes to the international literature on academic resilience by providing updated evidence from Italy within a comparative European perspective. The findings identify perseverance as the only socio–emotional skill consistently associated with academic resilience across Reading, Mathematics and Science, highlighting its central role in enabling socio–economically disadvantaged students to achieve outstanding academic performance. The study also shows that other socio–emotional skills display more limited and discipline–specific associations with academic resilience.

These findings provide further evidence that strengthening perseverance may represent a promising direction for educational policies and school interventions aimed at fostering academic resilience among socio–economically disadvantaged students.

Keywords: Academic resilience; Equity in education; Socio–emotional skills; OECD–PISA survey

Orchestrating Success: Music Proficiency, Emotional Intelligence and Mental Health in Young Adulthood

Lucia Mangiavacchi – Luca Piccoli – Giulia Gambardella

Introduction

Non–cognitive skills are increasingly recognized as central determinants of long–term educational and economic success (Deming 2017; Algan et al. 2022; Sorrenti et al. 2025). While school–based interventions targeting these skills have been studied extensively (Alan and Ertac 2018; Alan et al. 2019; Biroli et al. 2024; Alan and Kubilay 2025), the role of informal, leisure–based education, particularly musical training, remains comparatively underexplored. This paper estimates the causal effect of early–life musical training on emotional intelligence (EI) and mental health in young adulthood.

The existing evidence linking music training to non–cognitive development is largely correlational and vulnerable to selection: more talented and higher–SES children are more likely to take up music, and academically able children may self–select into musical practice (Hille and Schupp 2015; Cabane et al. 2016). We address this challenge by exploiting the staggered rollout of *Orchestra in Classe* (OIC), a universal, publicly funded orchestral–training program embedded in the primary–school curriculum of a municipality in Tuscany (Italy). OIC began in academic year 2007/08 in a single school before being extended to all primary schools in the municipality. Because enrollment was determined by school and cohort rather than by parental choice, participation is plausibly exogenous to family background and child ability, allowing us to use it as an instrument for sustained musical engagement during adolescence.

We do not evaluate OIC's direct effect on outcomes; instead, we model the program as the first step of a two–stage mechanism, whereby early exposure raises the likelihood that a child continues practicing music throughout adolescence. It is this sustained, endogenous engagement – observable now that the earliest cohorts have reached young adulthood – that we hypothesize shapes socio–emotional development. The transition to adulthood is itself a period of substantial change in educational, occupational, and social roles (Wachter 2020), making EI and mental health especially consequential outcomes at this stage of the life cycle.

Conceptually, active musical engagement is expected to strengthen self–regulation and goal–directed behavior (Uhlig et al. 2013), provide a structured outlet for emotional expression, and, through ensemble practice, build cooperative and relational skills. Music is also proposed as a protective factor against depression and anxiety through its capacity to modulate physiological stress responses (Schellenberg and Lima 2024). Building on the trait–EI framework of Salovey and Mayer (1990), our key contribution to the economics literature is to disaggregate emotional intelligence into its constituent subdomains, rather than treat non–cognitive skill as a single latent factor, in order to identify which specific facets of socio–emotional functioning respond to musical training.

Data used

We collected primary data through a web–based questionnaire administered between March and August 2024 to individuals who attended first grade in one of the municipality's five primary schools between academic years 2004/05 and 2009/10, a window spanning the introduction of OIC in the central school in 2007/08. The questionnaire covered family background, time use, and validated psychometric scales, and was linked to administrative records from the schools and the local music institute. To increase the response rate, we offered a participation incentive based on Induced Value Theory (Smith 1976). Using a snowball sampling procedure, we obtained 236 responses, 198 of which were fully completed, corresponding to roughly 31% of the target population of 636 children identified from school registries.

Comparisons with school registries (gender, cohort, school attended) and external sources (the Bank of Italy's Survey on Household Income and Wealth, ISTAT university enrollment statistics, and ESPAD–Italy) show that the sample is broadly representative, with a modest over–representation of females, the central (OIC) school, and middle–income households.

The two outcomes of interest are emotional intelligence, measured with the 30–item Italian–validated TEIQue–FF scale (Petrides 2009; Chirumbolo et al. 2019), and poor mental health, measured with the Italian–validated DASS–21 scale (Bottesi et al. 2015), covering depression, anxiety, and stress. For the main analysis we extract, via factor analysis, a single summary factor for each construct (eigenvalues of 7.23 and 8.98, respectively); we also analyze every subdomain of both scales separately. Control variables capture childhood family characteristics: gender, parental education, number of siblings, housing type, parental occupational skill, immigrant parentage, and family income (six brackets).

Methodology

Music proficiency. The survey collects twelve facets of musical practice (ability to play, number and type of instruments, years and weekly hours of practice, proficiency level, orchestra/ensemble participation). A factor analysis of these items yields a single dominant factor (eigenvalue 6.21; Cronbach's $\alpha = 0.7$), reported in Table 1, which we label *music proficiency*. Given its highly skewed, mass–point distribution (concentrated at a single value for non–musicians), our main specification uses a 0–10 rank transformation of this factor; the continuous factor and a proficiency dummy serve as robustness checks.

Table 1: Factor loading correlations and Eigenvalue for music proficiency

Musical practice variables	Factor 1	Uniqueness
Able to play a musical instrument	0.96	–0.07
Number of instruments played	0.97	–0.15
Years of musical practice (overall)	0.89	0.09
Maximum musical proficiency level for an instrument	0.92	0.04
Weekly hours of musical practice	0.62	0.34
Years at musical institute classes	0.59	0.46
Played in an orchestra, ensemble, band, etc.	0.89	0.11
Play keyboards (piano and similar)	0.63	0.09

Musical practice variables	Factor 1	Uniqueness
Play strings (guitar, bass guitar, etc)	0.41	0.15
Play winds (clarinet, flute, oboe, etc)	0.51	0.10
Play arch (violin, cello, viola, bass, etc)	0.48	0.30
Play brass (trumpet, sax, trombone, etc)	0.34	0.40
Eigenvalue	6.21	

Identification. Self–selection and reverse causality threaten OLS estimates of the music–EI/mental–health relationship: higher–SES families are more likely to pursue costly music education, and academically successful children may be more likely to take it up. We instrument music proficiency with OIC participation, exploiting its unanticipated, free, in–school, cohort–based rollout starting in a single school in 2007/08. Balance tests on pre–program school records show that OIC and comparison–school pupils are statistically indistinguishable in academic achievement and most family characteristics, with parental education the only notable imbalance, for which we control throughout. By contrast, a direct comparison of music–proficient and non–proficient respondents reveals large, significant gaps in academic achievement and socioeconomic status, confirming the scale of the selection problem that the instrument is designed to address.

Estimation. We estimate instrumental–variable models using both a control–function approach (Wooldridge 2015) and a Generalized Method of Moments estimator (Wooldridge 2010), clustering standard errors at the school–by–cohort level (47 clusters). To study which specific subdomains respond to music training, we estimate an IV Seemingly Unrelated Regressions (SUR) model via the control–function approach, recovering the effect of music proficiency on every TEIQue–FF and DASS–21 subdomain within a single system that allows arbitrary cross–equation error correlation. We further interact instrumented music proficiency with gender and parental immigrant status to assess heterogeneity. Because not all OIC participants became proficient musicians, IV coefficients are interpreted as Local Average Treatment Effects for the population of compliers.

Results and discussion

The first–stage relationship between OIC participation and music proficiency is strong (F–statistic of 61.3), mitigating concerns about weak–instrument bias. OLS estimates show no significant association between music proficiency and either outcome, consistent with offsetting selection and reverse–causality biases. The IV estimates, by contrast, are large and highly significant (Table 2): a one–point increase on the 10–point proficiency scale raises emotional intelligence by about 0.15 standard deviations and lowers poor mental health by about 0.11–0.12 standard deviations. The control–function and GMM estimators yield closely matching coefficients, and the significant first–stage residual confirms the endogeneity of music proficiency.

Table 2: IV estimates of the effect of music proficiency (0–10 rank)

Outcome	OLS	IV (Control function)	IV (GMM)
Emotional intelligence	0.016 (0.022)	0.146*** (0.041)	0.142*** (0.048)
Poor mental health	−0.000 (0.021)	−0.113*** (0.031)	−0.116*** (0.040)

Notes: Standard errors clustered at school/cohort level (47 clusters) in parentheses. First–stage F–statistic = 61.3. *** $p < 0.01$.

The subdomain (IV–SUR) analysis shows that the effect on emotional intelligence is concentrated in self–motivation, optimism, happiness, adaptability, assertiveness, low impulsiveness, and the quality of relationships, with smaller but significant effects on emotion regulation and emotional awareness; effects on emotion perception, expression, management, empathy, self–esteem, and stress

management are not statistically significant. For mental health, all three DASS–21 components decline significantly, most strongly depression and anxiety. This pattern points to a behavioral mechanism centered on self–regulation, reward, and ensemble–based relational skills, rather than on emotional perception or expression – consistent with, and providing causal corroboration for, a largely correlational psychological literature (Gómez–Zapata et al. 2021; Schellenberg and Lima 2024).

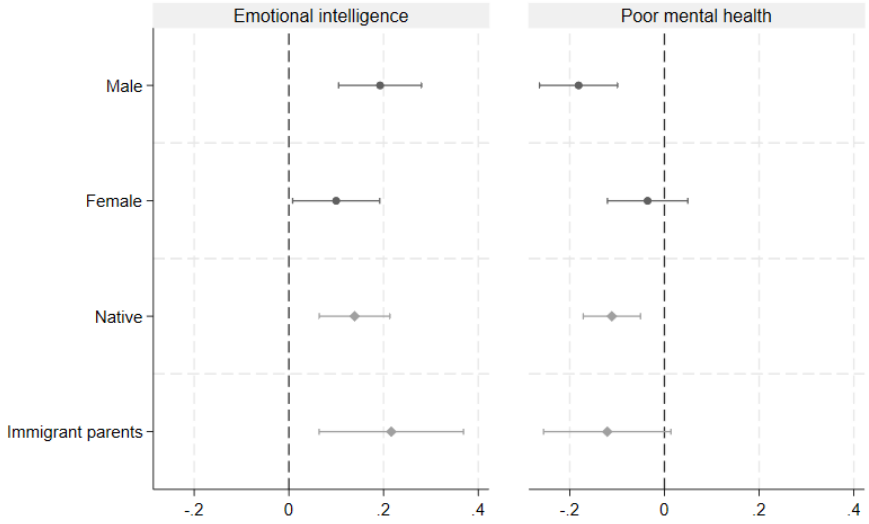


Figure 1: Differential impact of music rank by gender and parental immigrant status

Heterogeneity analysis (Figure 1) shows broader and larger effects for males, who display significant impacts across stress management, low impulsiveness, and all three mental–health dimensions, and for second–generation migrants, whose effects are particularly strong on relational and integration–relevant traits (relationships, emotion regulation, self–motivation, happiness, optimism, depression). While subgroup differences are not statistically distinguishable given the sample size, the pattern is consistent with music training compensating for group–specific socio–emotional barriers.

Conclusions

This study provides novel quasi–experimental causal evidence that sustained adolescent musical engagement, instrumented by participation in a universal school–orchestra program, improves emotional intelligence and reduces depression, anxiety, and stress symptoms in young adulthood. The concentration of effects in self–motivation, optimism, and relational competencies, together with the reduction in internalizing mental–health symptoms, suggests a behavioral mechanism rooted in the disciplined, goal–directed, and socially cooperative nature of musical practice rather than in its more passive, perceptual dimensions. Effects appear more pervasive for males and larger in magnitude for second–generation migrants, suggesting that music training may be particularly valuable for groups facing specific socio–emotional or integration barriers.

These findings support a policy rationale for publicly funded, universal music education as a cost–effective complement to cognitive–skill–oriented curricula, with value for both general human–capital formation and targeted integration policy. The main limitation of the study is sample size: while sufficient for the causal design employed, it constrains external validity. Future research should replicate this design in larger and more diverse settings and further trace the link between musical practice, academic achievement, and longer–run life–cycle outcomes such as educational track choice and university enrollment.

Keywords: Music training; Emotional intelligence; Mental health;Extracurricular activities

Socio–emotional factors as predictors of Implicit School Dropout. From the conceptual framework to guidelines for primary schools

Eugenia Cognigni

Introduction

Implicit school dropout refers to the condition of students who, despite completing their schooling, do not reach the expected minimum competence levels (Ricci, 2019): 12.3% of students are at risk at the end of lower secondary school (INVALSI, 2025, p. 166), while the rate of overt dropout at the end of upper secondary school stands at 8.7% (p. 169), with a significantly higher risk among students from disadvantaged socio–economic backgrounds (p. 168).

Early detection is strategic: risk indicators are already identifiable in primary school (Baldassarre & Sasanelli, 2020; Ricci, 2019; Arioli, 2025). Evidence documents the association between social–emotional competencies, academic outcomes, and educational equity (Cunha et al., 2010; Durlak et al., 2011; Kautz et al., 2014; Heckman, 2006; Dumfart & Neubauer, 2016; Honicke & Broadbent, 2016; OECD, 2021a, 2021b). Inclusive teaching practices create the conditions under which social–emotional competencies develop and, through them, sustain cognitive outcomes as well (Longobardi & Falzetti, 2025, pp. 93, 128, 130; Durlak et al., 2011; Kautz et al., 2014).

In the Italian context, however, this evidence has not yet led to a systematic understanding of which social–emotional factors may anticipate academic vulnerability and how they manifest themselves in everyday teaching practice. In this direction, Law No. 22 of 19 February 2025 introduces the development of non–cognitive and transversal competencies across all levels of the education system. It launches a three–year national pilot programme to identify methodologies, good practices, and assessment tools, among whose stated aims are reducing school dropout, including its implicit form, and preventing functional illiteracy.

The theoretical taxonomy adopted in this research draws on three complementary frameworks: CASEL (2020), the OECD Survey of Social and Emotional Skills (OECD, 2021, 2025), and ENRICH–INVALSI (Longobardi & Falzetti, 2025). The latter represents a first step towards addressing a significant knowledge gap in the Italian context, where no systematic nationwide assessment of students' social–emotional competencies is currently available (Longobardi & Falzetti, 2025, pp. 48–49).

Building on this theoretical framework, the study pursues five objectives: to investigate whether the social–emotional factors associated with the domains of the adopted taxonomy emerge as early signals of academic vulnerability in primary school classrooms, as reflected in teachers' representations and practices; to verify whether teachers systematically recognise the social–emotional factors included in the adopted taxonomy; to describe whether schools' socio–economic profile influences teachers' ability to identify these factors; to describe whether inclusive teaching practices foster their development; and to verify whether teachers confirm the adopted taxonomy while identifying its specific features within the primary school context.

Data Used

The research selects Lombardy within the North–West macro–area, which accounts for over 1,115 state schools (MIM, 2024). The macro–area presents a "geografia capovolta" ("reversed geography") (Longobardi & Falzetti, 2025, p. 97): despite recording some of the lowest rates of implicit school dropout at the national level (3.7% at the end of upper secondary school, compared with a national average of 8.7%; 7.6% of students at risk of implicit dropout at the end of lower secondary school, compared with a national average of 12.3%; INVALSI, 2025), the North–West reports social–emotional scores below the national average across almost all competencies (Longobardi & Falzetti, 2025, pp. 93–99). This condition makes it possible to observe latent risk before it develops into manifest implicit school dropout.

INVALSI supported the selection of four Comprehensive Institutes (Istituti Comprensivi) through purposive sampling. Two criteria guided the selection: the school's socio–economic profile (ESCS index, used as a proxy for the socio–economic context and measured through a student questionnaire administered at the end of Grade 5 in the 2023/24 school year) and students' performance in the INVALSI test outcomes (high scores in the Grade 2 primary school tests, 2023/24 school year). The intersection of these two criteria defines an embedded multiple–case design with four cells: two focal cases (high INVALSI–low ESCS; high INVALSI–high ESCS), in which the study investigates the presence of latent forms of implicit school dropout under conditions of socio–economic disadvantage and advantage, respectively; and two contrasting cases (low INVALSI–low ESCS; low INVALSI–high ESCS), which make it possible to examine whether the patterns identified in the focal cases depend on socio–economic background alone or on academic performance alone (Trochim, 1989). The study adopts a longitudinal design. It follows the same cohort of pupils over time, from the selection of schools based on Grade 2 INVALSI outcomes (2023/24 school year) to the beginning of fieldwork with the same cohort in Grade 5 (2026/27 school year). The overall sample consists of 4 Comprehensive Institutes, 12 Grade 5 primary school classes, and 24 teachers.

Methodology

The present research adopts an exploratory–interpretive qualitative paradigm (Lincoln & Guba, 1985). Social–emotional factors emerge through situated behaviours and representations that require contextual understanding before they can be translated into measurable indicators (Lincoln & Guba, 1985; Yin, 2018). A multiple comparative case study design (Yin, 2018) makes it possible to explore the situation within its natural context while preserving the complexity of the relationships among social–emotional factors, teaching practices, and socio–economic context. Comparing the four cases makes it possible to isolate the contribution of the socio–economic variable (Yin, 2018). The findings are not statistically generalisable but transferable to analogous contexts (Lincoln & Guba, 1985).

The study collects data through three complementary techniques carried out in parallel. Classroom observation, conducted both by the classroom teacher and by the doctoral researcher, provides two perspectives on everyday teaching practice and is supported by a semi–structured observation form completed independently by both at the end of each lesson. Semi–structured interviews with teachers explore their representations, intentions, and interpretative reasoning underlying the observed practices. Focus groups provide a space for collective discussion among teachers, generating shared meanings as well as different perspectives on social–emotional factors. The use of multiple sources of evidence enables triangulation: comparing evidence from classroom observations, the semi–structured instrument, and qualitative textual data strengthens the credibility of the findings and highlights the complexity (Lincoln & Guba, 1985; Nowell et al., 2017).

The adopted taxonomy includes four domains (Task Performance, Open–Mindedness, Collaboration, and Emotional Regulation) and seven competencies (Persistence, Responsibility, Curiosity, Empathy, Stress Resistance, Self–Control, and Emotional Control). The selection follows two explicit criteria. First, each competency had to be predictive of academic success, that is, among the strongest predictors of learning outcomes identified in the literature. Second, each competency had to show comparatively lower scores in the North–West macro–area, namely competencies for which the ENRICH survey reports scores below the national average. Within this framework, the study intentionally includes Self–Control and Emotional Control in the taxonomy. The ENRICH survey excludes these competencies as a precautionary methodological choice related to participants' age, the sensitivity of the topics addressed, and the need to maintain an appropriate approach to the school context.

The research design includes a semi–structured observation form which, once validated, is intended to answer the following research question: Are social–emotional factors recognised by teachers in everyday classroom practice? (RQS1). Grade 5 primary school teachers complete the form at the end

of each lesson and is consists of four sections: Part A (Detecting), Part B (Recognising), Part C (Describing and Interpreting), and Part D (Contextual Notes) (Jacobs et al., 2010; Flanagan, 1954; Tripp, 1993; Castoldi, 2009; Dewey, 1910; Schön, 1983; Mortari, 2003).

Part A consists of items corresponding to the seven competencies listed above. These items were adapted and modified from the Student Questionnaire developed for the OECD Survey of Social and Emotional Skills (SSES 2019), following formal authorisation from the OECD. The construction of the teacher observation instrument also draws on the definitions of social–emotional competencies provided in Beyond Academic Learning: First Results from the Survey of Social and Emotional Skills (OECD, 2021a, p. 34) and in the SSES 2019 Technical Report (OECD, 2021b, p. 17). The adaptation process takes participants' age and the specific characteristics of the primary school context into account. Before its use in the field, the observation form undergoes a three–phase validation process. The validation is not psychometric in nature but aims to examine the theoretical coherence and practical applicability of the instrument, consistently with the qualitative design of the study. Its purpose is to refine the observation form before field implementation. The validation process involves 3 Comprehensive Institutes, 14 teachers, six Grade 5 classes, and 24 hours of classroom observation. Part A employs five–point rating scales that generate ordinal quantitative data. These data are analysed through descriptive statistics, including frequency distributions, medians, and modes, together with a comparative descriptive analysis across indicators and domains. Parts B and C collect qualitative textual data through open–ended responses. These data are analysed using Reflexive Thematic Analysis (Braun & Clarke, 2006) through the development of codes and themes. Integrating the two levels of analysis enables a comparison between what Part A detects — the perceived prevalence of social–emotional factors at the classroom level — and how teachers recognise and interpret those factors in Parts B and C.

The study pursues trustworthiness according to the four criteria proposed by Lincoln and Guba (1985) and further discussed by Nowell et al. (2017). The study strengthens credibility through triangulation and member checking. It supports transferability through thick description of the research context. It ensures dependability through an audit trail, systematic tracking procedures, and consistency of coding over time. Finally, it ensures confirmability by grounding interpretations in the empirical data rather than in prior assumptions, through the doctoral researcher's reflexive positioning and explicit clarification of meaning.

Ethical considerations inform every stage of the research process. The study guarantees participants' anonymity, secure data storage in compliance with the General Data Protection Regulation (EU Regulation 2016/679), and a faithful and respectful representation of participants when speaking of, with, and for them.

Results and Discussion

The study produces two complementary outputs. Output 1 – Conceptual framework. The study develops a conceptual framework that will provide the basis for the future development of multidimensional early warning systems for Italian primary schools, in collaboration with the University of Twente and CITO (The Netherlands). Output 2 – Operational guidelines. The study will co–construct a set of operational guidelines with teachers to support the recognition of social–emotional factors as early signals of the risk of implicit school dropout. The guidelines support primary school teachers and school leaders in recognising social–emotional factors. Their co–construction reflects a shared process of knowledge building (Wenger, 1998; Penuel et al., 2015). The conceptual framework provides the theoretical structure for the assessment process, whereas the operational guidelines translate that framework into everyday teaching practice.

Several limitations accompany the expected findings, and the research design explicitly addresses them. First, the data depend on teachers' ability to recognise and describe social–emotional factors, as they are the research's privileged informants. Second, the presence of an external observer in the classroom may influence participants' behaviour through the Hawthorne effect. To minimise this risk, the doctoral researcher adopts a discreet and non–intrusive observational stance throughout

the fieldwork (Trinchero, 2004, p. 118). The study does not aim to produce statistically generalisable. Rather, the study aims to formulate «statements of possibility», identify recurring regularities across cases, and develop a coherent conceptual framework (Trinchero, 2004, p. 8). Detailed description and interpretation of the research context will enable identification of analogous settings to which the findings may be transferred. Furthermore, the explicit account of the epistemological and methodological choices underpinning the study — including the doctoral researcher's positioning in the field — will enable other researchers to retrace the research process under comparable conditions and reach compatible interpretations. These elements support the validity of the findings within a qualitative research perspective. The research therefore treats the absence of experimental control over variables not as a limitation but as an epistemological choice, consistent with an interpretive approach aimed at understanding educational reality in its situated complexity (Yin, 2018; Lincoln & Guba, 1985; Trinchero, 2004). The research values participants' perspectives as the basis for interpreting the reality under investigation. It establishes a dialogue with educational practice by producing knowledge intended to inform and improve teachers' professional work. The study directs attention towards individual cases and their specific characteristics: the study interprets empirical evidence to generate knowledge grounded in the educational reality under investigation. Consequently, no rigid, predetermined framework guides data collection. Rather, participating teachers identify the information that is most meaningful in relation to the research objectives. Finally, the doctoral researcher's values, assumptions, and subjectivity may represent a potential source of bias. The research addresses this risk through continuous reflexivity, regular discussion with the research supervisors (tutor and co–supervisor), and ongoing dialogue with other researchers acting as expert discussants in seminar settings (Viganò, 2020; Trinchero, 2004).

Conclusions

At the current stage, the study has led to the development of the theoretical framework, the identification of the criteria for school selection, and the ongoing validation of the observation form. The research will complete the empirical phase during the first half of next year before moving to the co–construction of the operational guidelines with participating teachers. The study will determine whether the social–emotional factors, identified in the adopted taxonomy, constitute early signals of academic vulnerability that teachers can recognise in everyday classroom practice. From a theoretical perspective, the study proposes an exploratory interpretive framework that clarifies the relationship between social–emotional competencies and academic vulnerability, addressing a research area that remains underexplored at the intersection of the literature on implicit school dropout and social–emotional development in primary education. From a practical perspective, the study provides evidence to inform teacher professional development and the integration of social–emotional dimensions into early warning systems.

Keywords: Implicit school dropout; Social–emotional predictive factors; Teachability; Primary school

Civic Agency in the Diagnosis Phase of Service–Learning Pathways: Evidence from Italian Schooling and Implications for Assessing Social and Emotional Skills

Simone Consegnati – Patrizia Lotti

Introduction

The diagnosis phase, understood as the stage in which students, teachers, and the community jointly investigate needs and build a critical reading of the context, is a decisive step in Service–Learning (SL) pathways. Despite its pedagogical relevance, this phase remains under–documented in the research literature, especially in relation to primary and secondary schooling. This contribution examines that phase across the full vertical curriculum of the Italian school system, with the aim of understanding to what extent, and through which practices, schools are already fostering forms of civic agency that, in the panel's framework, can be closely linked to social and emotional learning (SEL) dimensions such as collaboration, responsibility, listening, participation, and the ability to read real–world situations reflectively.

The study is framed within SL as an approach that integrates curriculum learning, civic participation, and engagement with the local territory. From this standpoint, examining how schools structure the diagnosis phase means asking what educational conditions prepare students for more mature forms of participation, shared responsibility, and co–design with community partners. This perspective is consistent with the session's focus on the development and assessment of social and emotional skills through methodologically grounded approaches.

The literature supports this assumption: the quality of the initial phase shapes the effectiveness of the entire pathway, distinguishing an SL model that treats the community as a mere recipient from one that involves it as a co–constructor of needs from the diagnosis stage itself (Zeichner, 2009). Implementation fidelity, often linked to structured assessment devices that are largely absent here, is a critical factor for outcomes (Kim, Burkhauser, & Quinn, 2017; O'Donnell, 2008), outcomes that the meta–analysis by Celio, Durlak, and Dymnicki (2011) associates positively with SL. Even more experienced teachers, according to Wolff, van den Bogert, and Jarodzka (2014), tend to shift attention toward the external context—a movement that remains relatively uncommon in our data. The taxonomic reference remains the five–dimension model proposed by Weissberg and colleagues (2015), taken up by the CASEL framework (2020) and operationalized internationally by the OECD (Kankaraš & Suárez–Álvarez, 2019); since self–report instruments remain vulnerable to bias (Duckworth & Yeager, 2015), triangulation with observational and behavioral measures (Kankaraš, Feron, & Renbarger, 2019) offers the most appropriate framework for situating the practices analyzed here.

The contribution's relevance to the panel lies mainly in the diagnostic phase as an observable setting in which competencies that are often difficult to capture through self–report measures take shape in concrete practices. Involving students in identifying needs, defining problems, and engaging with external actors makes it possible to observe situated indicators of agency, cooperation, contextual sensitivity, and responsible decision–making. The paper therefore does not attempt to redefine the original dataset, but rather rereads it through an interpretive lens that highlights the link between SL practices and the development of social and emotional competencies in school settings.

Data used

The study is based on a quantitative survey conducted by INDIRE within the Avanguardie Educative movement, Italy's largest school innovation network. Of the 645 schools with SL experience, 453 had designated a SL contact teacher by adopting the idea "Inside/Outside the School – Service–Learning" or by taking part in the annual survey on the institutionalization of SL; these teachers received a structured questionnaire by email. Twenty–seven emails bounced, leaving 426 successfully delivered. A total of 176 responses were received, coming from 108 schools, for a response rate of

41.3% calculated on delivered emails (176/426). The responding schools cover the entire vertical curriculum: early childhood education (13%), primary school (40%), lower secondary school (20%), and upper secondary school (27%). The questionnaire items capture specific aspects of the diagnosis phase in SL pathways. They concern: the degree to which the phase is structured; the role assigned to students; the teaching strategies adopted; the involvement of community partners; documentation practices; and assessment practices. The dataset therefore makes it possible to observe not only whether the diagnosis phase is present, but also the organizational and pedagogical forms through which it is enacted across different school levels.

Methodology

The study adopts a descriptive quantitative design based on a structured questionnaire. The analysis aims to reconstruct how schools organize the diagnosis phase in SL pathways and to highlight possible differences across school levels. Five analytical areas are considered: student participation; teacher role; involvement of community partners; documentation practices; and assessment practices. The data are read from a vertical perspective in order to capture continuity and differences along the school pathway. This design makes it possible to examine the extent to which schools provide progressive opportunities for students to exercise agency and assume responsibility. For the purposes of the panel, the findings are also discussed as indicators of contexts that foster social and emotional skills, especially when the diagnosis phase prompts contextual observation, reciprocal listening, interaction with the territory, and shared decision–making. Methodologically, the contribution takes a cautious stance. It does not automatically equate civic agency with social and emotional skills, nor does it overstate what the data can show. Instead, it proposes an analytical reading in which some practices associated with the diagnosis phase are treated as educational conditions or observable proxies of relevant social and emotional processes. In this sense, the paper speaks directly to the panel by underscoring the need to combine multiple tools and sources in order to capture complex competencies that emerge through behavior, relationships, and authentic learning situations.

Results and discussion

The findings show widespread attention to student agency from the very beginning of the diagnosis phase. Seventy–one percent of teachers report involving students from the outset, a figure that rises to 85% in upper secondary school. Teacher roles also tend, especially at this level, to be non–directive: when the two response options describing a posture of shared co–exploration and facilitation of students' independent inquiry are combined, this orientation reaches 70.5% overall and 75% in upper secondary school. A second important result concerns openness to the outside world. Early involvement of community partners is reported in 45% of cases and reaches 52% in upper secondary school. Systematic documentation of the phase is also more established at higher school levels: the overall figure is 75%, rising to 88% in upper secondary school. These findings suggest that, as students grow older, they are more likely to take part in richer practices of contextual analysis, interaction with social actors, and shared problem construction. A complementary reading emerges from the cluster analysis conducted on the same data (K–means, k=4). The four identified profiles differ sharply in their openness to the territory, student roles, and documentation and assessment practices. What makes this reading especially relevant, however, is a transversal finding: none of the four profiles shows the involvement of external actors in the diagnosis phase in a strict sense. If one accepts that co–constructing needs with external actors fosters dimensions such as social awareness and responsible decision–making, this pattern points to a structural limitation of the system, independent of school level and implementation profile, rather than to a lack of pedagogical sensitivity among individual teachers.

Important cross–cutting weaknesses remain across school levels. Seventy–two percent of teachers report working without a shared institutional framework, and a significant share indicate that learning developed during the diagnosis phase is not formally assessed. This is especially important in the context of the panel because it reveals a methodological and policy problem at once:

educational practices that may foster relevant competencies often remain implicit, weakly formalized, and difficult to translate into evidence that can inform school improvement and policy design.
Three points follow from this discussion. First, the diagnosis phase in SL appears to be a promising context for observing competencies with a strong social and emotional component, even when these do not surface well through self–report tools. Second, the fact that many schools activate participation, exploration, and collaboration without a corresponding assessment framework suggests the existence of a widespread but largely invisible educational capital. Third, the gradual increase in student involvement, teacher facilitation, and community openness along the curriculum points to a developmental pathway that could be investigated further through more integrated designs combining descriptive data, structured observations, assessment rubrics, and outcome measures.

Conclusions

The contribution shows that the diagnosis phase in SL pathways is a pedagogical laboratory of particular interest for the study of social and emotional competencies. It makes observable processes such as active participation, shared responsibility, engagement with different viewpoints, the ability to read real–world problems, and co–construction of responses with others. These processes do not exhaust the field of social and emotional skills, but they do capture important dimensions in concrete, situated, and educationally meaningful ways.
The analysis also shows that this potential remains partly unrealized, because many practices are not supported by shared institutional frameworks, common documentation tools, or systematic assessment procedures. This creates a central tension for research and school policy alike: schools already seem to offer meaningful opportunities for the development of transversal competencies, yet their measurement and formal recognition remain weak. In line with the session's aims, the paper therefore suggests the need for more robust and context–sensitive assessment tools that can recognize what happens in authentic learning and participation pathways.
This perspective opens two lines of work. On the research side, it invites closer connections between studies of SL practices and the broader debate on assessing social and emotional competencies through multi–method approaches. On the policy side, it suggests the value of treating the diagnosis phase not only as preparation for service action, but also as a formative space in which crucial competencies for democratic citizenship and relational well–being are built.

Keywords: Service–Learning; civic agency; Social and emotional skills; Assessment; Multi–method assessment

Embedding Socio–Emotional Skills into Everyday Teaching: Preliminary Findings from an Experimental Pilot in an Italian School

Giulia Vaiani – Gabriele Iannotta – Tommaso Agasisti

Introduction

In the last fifteen years, social and emotional learning (SEL) has become a central topic in education research and policy. Social and emotional skills (SES) –such as self–awareness, self–management, social awareness, relationship skills, and responsible decision–making– are increasingly recognized as key drivers of students' well–being, school climate, and long–term life outcomes. This growing interest is reflected in the large number of school–based SEL programs implemented worldwide, as well as in a rapidly expanding body of meta–analytic evidence showing that these programs produce significant, positive effects on students' skills, attitudes, and behaviors.

Despite this growing evidence base, an open question in the literature concerns how SES should be developed in schools. One approach treats SEL as a stand–alone subject, with dedicated lessons and a separate curriculum; this is, in fact, the dominant model among the interventions evaluated so far, and it underlies most of the documented positive effects. An alternative approach embeds SES development directly into ordinary subject teaching, through the way lessons are delivered, rather than through additional content. This approach has practical advantages that may matter for adoption and sustainability: it requires no additional teaching time and can be more easily embedded in an already demanding curriculum, which motivates its use in the present study.

A further reason to study SES is that they are not only valuable in themselves, but have also been shown to be associated with more traditional academic outcomes, such as grades, test scores, and school functioning. This suggests that investments in SES may generate returns that go beyond well–being, also supporting learning outcomes typically targeted by educational policy.

Objective of the study

This paper presents preliminary evidence from a pilot experimental study conducted in a single private K–12 school in Milan, Italy. The school serves a socio–economically advantaged student population, which represents a specific and somewhat unusual context compared to the majority of the SEL literature, but also offers a controlled environment in which to test the feasibility and initial effects of an intervention based on non–traditional teaching practices.

The study combined a qualitative preparatory phase with a quantitative impact evaluation. The main objective of the study is to measure the effect on students' socio–emotional skills of lessons delivered through non–frontal, more participative teaching methods (e.g., group work, peer collaboration, project–based activities), compared to standard teaching practices, without introducing separate SEL lessons or reducing curricular teaching time.

Empirical Strategy and Data Collection

The project started with a qualitative phase aimed at understanding the school's context and needs. The qualitative phase began with an interview with school leadership, followed by semi–structured interviews with 13 teachers, selected to reflect diversity in seniority, professional background, school level (from early childhood to upper–secondary education), subject taught, and role, and including at least one special–education teacher. Interviews were conducted in person over four sessions within about two weeks, and lasted between 30 and 45 minutes each. These interviews aimed to understand how socio–emotional competences were already perceived and practiced across the school, and – specifically among teachers from the primary and lower–secondary levels targeted by the pilot intervention– to identify a small set of priority socio–emotional competences for each of these two school levels. Classroom observations of teachers who already used innovative, non–frontal teaching approaches were also carried out, to document good practices already present in the school.

Based on these interviews, three priority socio–emotional competences were identified for each school level. For primary school, these were: teamwork and collaborative problem–solving, effective communication, and stress–management strategies. For lower–secondary school, these were: emotion regulation, teamwork and collaborative problem–solving, and self–efficacy.

A group of teachers was then engaged in a multi–month training program on SEL. The teacher training program was delivered by an external partner organization specialized in emotional intelligence education, through a multi–month certification course combining theoretical content, experiential activities, and a final individual or group project applying the training content to classroom practice. The priority socio–emotional competences targeted by the intervention, as well as the student and teacher questionnaires, were instead defined and measured following the CASEL framework and an Italian adaptation of the OECD Survey on Social and Emotional Skills, respectively.

The impact evaluation follows an experimental, matched treatment/control design at the class level. Fifth–grade classes were assigned to treatment and fourth–grade classes to control in primary school; second–grade classes were assigned to treatment and third–grade classes to control in lower–secondary school. Grade levels were chosen based on considerations of group stability and, in lower–

secondary school, on the fact that final–year classes are strongly exam–oriented while first–year classes often lack a consolidated group identity. Within each pair of grades, teachers were assigned to treatment or control classes so as to avoid any teacher teaching in both conditions. Students in treatment classes were taught by teachers who had completed the training program and were asked to apply non–frontal, collaborative teaching practices in their ordinary lessons; students in control classes continued with standard teaching practices.

Data collection relies on student self–report questionnaires, based on the Italian translation of the OECD Survey on Social and Emotional Skills and complemented with additional items targeting the three priority competences identified for each school level. Beyond these priority competences, the instrument also captures broader dimensions of student well–being, including life satisfaction, perceived safety, peer and teacher relationships, bullying experiences, and students' mindset about their own social, emotional, and cognitive abilities; item wording was adapted separately for primary and lower–secondary school students. A shorter, teacher–administered instrument was collected in parallel, in which teachers rated each student on the same three priority competences using a four–point ordinal scale.

Both instruments were administered before the start of the intervention (baseline) and then again after the intervention period, roughly three months later, to allow for a pre/post comparison between treatment and control classes. The baseline data collection covered 323 students (120 in primary school and 203 in lower–secondary school), across seven primary school classes and fourteen lower–secondary school classes assigned to the treatment or control condition. A smaller number of students could be matched across the pre– and post–intervention waves, reflecting some attrition between the two data collection rounds; the extent and balance of this attrition across treatment and control groups will be assessed in the full paper.

Preliminary Results

At this stage, only baseline (pre–intervention) data have been analyzed; the post–intervention data collection and the estimation of treatment effects are still ongoing. The baseline descriptive analysis shows an overall positive picture in terms of students' life satisfaction, perceived safety, and relationships with peers and teachers, together with generally favorable levels of the targeted priority competences. At the same time, the baseline data also highlight some areas of attention, including perceived pressure from adults, occasional experiences of peer conflict, and a non–negligible share of students reporting a relatively fixed mindset about their social, emotional, and cognitive abilities. These baseline patterns provide a starting point against which post–intervention changes will be assessed, and they also reveal meaningful variation across classes within the same school, which will be taken into account in the impact evaluation.

Preliminary descriptive comparisons between treatment and control classes are currently being examined, and a more detailed analysis is underway. This analysis is expected to focus primarily on the lower–secondary school sample, since the number of classes assigned to each condition differs substantially between school levels. Indeed, while the lower–secondary sample includes a comparable number of treatment and control classes, the primary school control group consists of only two classes, which severely limits the robustness of any class–clustered comparison at that level. As a further extension, we plan to explore the relationship between changes in socio–emotional skills and students' performance on standardized INVALSI tests.

Full results, together with an assessment of baseline equivalence between treatment and control groups, will be presented in the completed paper.

Keywords: Socio–emotional skills (SES); School–based intervention; Teacher training; Experimental design

SESSION 1. EVALUATING THE IMPACT OF EDUCATION POLICIES: COUNTERFACTUAL APPROACHES AND EVIDENCE FOR BASIC SKILLS IMPROVEMENT

ORGANIZER: EUROPEAN COMMISSION
COORDINATORS: DANIELE VIDONI
4TH NOVEMBER: 4.30 P.M. – 6.30 P.M. {AULA MAGNA – RESEARCH 4}

The impact of congestion pricing policies on primary school students' performance: individual–level evidence from Milan's Area C

Léonard Moulin – Valeria Maria Urbano – Lorenzo Maraviglia

Introduction

Air pollution represents one of the most relevant environmental challenges in urban areas. A large body of evidence identifies vehicular traffic as one of the primary sources of air pollution in cities, with detrimental effects on human health, cognitive functioning, and economic productivity (Davis, 2008). To address these challenges, several European cities have introduced low emission zones (LEZs), designated urban areas in which access is restricted for vehicles that do not meet specific emission standards. A growing empirical literature shows that LEZs are effective in reducing traffic–related air pollution, generating significant environmental and health benefits. Given the wide spectrum of potential benefits and costs, it is crucial to explore the broader externalities associated with LEZs.

Compared to adults, children are particularly susceptible to environmental hazards, as their respiratory and immune systems are still developing. Recent research has, therefore, begun to explore not only the direct health consequences of LEZs for children health, but also their broader effects on educational achievement. Conte Keivabu and Rüttenauer (2022) explored the impact of stricter traffic regulations introduced in Central London and revealed significant reductions in absences among schools with a high share of students with low socio–economic status. Avila–Uribe et al. (2024) analyzed the introduction of the Ultra Low Emission Zone in London showing that improved air quality around schools led to an average increase of 0.09 standard deviations in standardized test scores. Brehm et al. (2025) found that LEZ implementation increased transition rates to the academic track in secondary education, and Valdés et al. (2025) documented that the LEZ introduced in Madrid generated a 0.20 standard deviation increase in university entrance exam performance. Despite providing consistent preliminary evidence, these studies rely on aggregated school–level outcomes. This limits the precision of estimates, prevents direct control for compositional changes in school enrollment over time, and does not allow for the analysis of how effects are distributed across students with different characteristics.

This paper examines the educational effects of Milan's Area C congestion policy using individual–level student data spanning academic years 2009/10 to 2018/19. Milan provides a particularly relevant setting for studying the broader externalities associated with the introduction of the LEZ. The city is Italy's financial and economic capital and one of the most polluted urban areas in Western Europe, partly because of its location in the Po Valley, where geographical and meteorological conditions favor the accumulation of pollutants. The contribution of this study is threefold. First, by examining the case of Milan's Area C, it extends the existing evidence to a new institutional and geographical context. Second, it provides the first individual–level analysis of the impact of a LEZ on academic achievement at the primary school level, enabling a direct assessment of how congestion reduction policies affect the distribution of learning outcomes across students, free from the

compositional confounders that can bias school–level estimates. Third, the paper provides a systematic comparison of effects across two grades within primary school (Grade 2 and Grade 5) and across two subjects (Mathematics and Italian), examining whether the benefits associated with improved air quality differ by developmental stage and by the cognitive demands of different academic disciplines.

Data

We rely on individual–level data from the INVALSI standardized assessments. Since the 2009/2010 school year, second– and fifth–grade students — the two primary–education cohorts assessed by INVALSI — have taken these standardized national tests on an annual basis. Participation of schools in these national tests is compulsory by law, which ensures full participation and limits selection bias in test–taking. The tests cover both Mathematics and Italian language. Scores used in the analysis consist of the percentage of correct answers and are standardized to have zero mean and unit variance within each school year. INVALSI data also provide individual–level information on children's and parental characteristics, including gender, origin, employment status of the parents, regularity of school progression, and attendance at pre–primary school. After excluding schools with missing geolocation and those without pre–treatment observations, the final analytical sample comprises 329 schools, of which 17 are located in Area C and 312 in the wider Milan Metropolitan Area.

Introduced on January 16, 2012, Area C combines access restrictions for the most polluting vehicles with a flat daily charge for vehicles allowed to enter, with the explicit objective of reducing both traffic congestion and atmospheric pollution. The policy generated a sharp and well–documented reduction in traffic and air pollution (Beria, 2016; Gibson and Carnovale, 2015; Percoco, 2013), exposing schools located within the charging zone to an abrupt improvement in their local environment. Using open data from the air pollution monitoring system of ARPA Lombardia, we document the evolution of particulate matter (PM_{10}) and nitrogen dioxide (NO_2), two pollutants commonly associated with vehicular traffic. The estimated differences on both pollutants are negative and statistically significant for all years considered, confirming previous empirical findings that air pollutant levels declined over time following the policy implementation.

Methodology

We estimate the impact of the introduction of Area C on student achievement using a two–way fixed effects (TWFE) difference–in–differences framework, exploiting variation across schools and over time:

$$Y_{ist} = \alpha + \beta(AreaC_s \times Post_t) + \boldsymbol{X}'_{ist}\gamma + \mu_s + \lambda_t + \varepsilon_{ist}$$

where Y_{ist} denotes the standardized INVALSI test score of student i enrolled in school s in year t. The treatment indicator $AreaC_s$ equals one for schools located within the Area C zone, while $Post_t$ is an indicator equal to one for years following the policy implementation in 2012. The coefficient of interest, β, captures the average effect of the policy on student performance. The vector $\mathbf{X}_{ist}$ includes a set of student–level covariates: parental occupation, gender, migration background, early childhood education attendance, and indicators for grade retention or irregular school trajectories. We include school fixed effects μ_s and year fixed effects λ_t, and we cluster standard errors at the school level.

Identification relies on the assumption that, absent the introduction of Area C, treated and control schools would have followed parallel trends in test scores. A test for differential linear trends between treated and control schools in the pre–treatment period yields an F–statistic of 0.01 (p–value = 0.91), failing to reject the null hypothesis of parallel trends. To further address potential concerns about differential pre–treatment dynamics, we complement the baseline estimates with the Synthetic Difference–in–Differences (SDID) estimator (Arkhangelsky et al., 2021), which reweights control schools so that their weighted average better reproduces the pre–treatment trajectory of treated schools. We next examine the dynamic effects of the policy using an event–study specification,

and we explore whether the impact of Area C varies across student characteristics by estimating an interaction model in which the treatment indicator is interacted with student–level characteristics.

Results and discussion

Across the full sample, the difference–in–differences estimates reveal positive and statistically significant effects for second–grade pupils, while the coefficients for fifth–grade students are consistently close to zero and not statistically significant (Table 1). For Grade 2, exposure to Area C improves both Mathematics and Italian performance: Mathematics scores increase by 0.256 standard deviations (s.d.) and Italian scores by 0.116 s.d. Analytical subjects, such as mathematics, typically require problem–solving and higher–order reasoning, whereas language–based subjects often rely more on memorization and lower cognitive effort; since more cognitively demanding tasks are likely to be more sensitive to environmental stressors, the higher effect on Mathematics is consistent with previous studies on air pollution (Duque and Gilraine, 2022). The pattern of sizeable effects in Grade 2 but no detectable impact in Grade 5 is consistent with the idea that air pollution reductions have greater impacts on younger children, in line with evidence that cognitive development is more malleable and, therefore, more vulnerable to environmental exposure at younger ages (Persico and Venator, 2021).

Table 1. Effect of Milan's Area C congestion policy on standardized test scores: baseline DID estimates

	Grade 2		Grade 5	
	Math	Italian	Math	Italian
Area C × Post–treatment period	0.256***	0.116*	0.142	0.073
	(0.085)	(0.063)	(0.115)	(0.080)
Year fixed effects	X	X	X	X
School fixed effects	X	X	X	X
Covariates	X	X	X	X
N	213,360	213,486	204,828	205,222

Notes: All specifications include year and school fixed effects and a rich set of student–level covariates (parental occupation, gender, migration background, early childhood attendance, grade regularity). Standard errors clustered at the school level in parentheses.

* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

The SDID estimates are broadly consistent with the baseline findings, although reduced in magnitude. For Grade 2 Mathematics, the SDID coefficient remains positive and statistically significant, with an estimated effect of 0.219 s.d., while for Grade 5 the point estimates are close to zero in both subjects. The dynamic patterns are consistent with the evolution of pollution levels following the introduction of Area C: the years in which we observe the strongest and most robust effects on student achievement in Grade 2 are precisely those in which the largest and most statistically significant reductions in both NO_2 and PM_{10} concentrations occur. This temporal alignment suggests that the gains in student performance emerge most clearly when reductions in air pollution are both substantial and statistically significant.

The heterogeneity analysis reveals that the gains from Area C are not uniform across students. The most pronounced differences arise along socio–economic dimensions: in particular, students whose parents hold executive occupations exhibit negative estimated effects, which contrasts with the positive average treatment effects and suggests relatively more favorable impacts for students from less advantaged backgrounds. The overall positive average impact implies relatively larger gains among students with less favorable initial conditions, a pattern consistent with a reduction in early achievement gaps. The magnitude of our estimates is within the range documented in the existing literature on environmental regulation and educational outcomes (Avila–Uribe et al., 2024; Valdés et

al., 2025). Our estimates capture the overall effect of Area C on student performance, but they do not allow us to isolate the specific mechanisms through which this effect operates: because Area C affected traffic, air pollution, and potentially noise at the same time, our estimates cannot separately identify the contribution of each environmental channel.

Conclusions

Our central finding is that the introduction of Area C generated significant improvements in early academic achievement, with effects concentrated among second–grade students. Exposure to the congestion policy leads to a statistically significant increase of 0.256 s.d. in Mathematics scores and 0.116 s.d. in Italian scores for Grade 2 students, and these results are robust across a wide range of alternative specifications, sample definitions, and inference methods. By contrast, no statistically significant effects emerge for fifth–grade students in either subject. The heterogeneity analysis points to an equity–enhancing dimension of congestion pricing that has received limited attention in the existing literature: beyond reducing traffic and improving air quality, such policies may generate significant redistributive effects on children's human capital formation, narrowing early inequalities that closely mirror the socioeconomic structure of the city. When assessing the welfare effects of such policies, standard cost–benefit analyses that focus exclusively on congestion relief, fuel savings, and health outcomes are likely to substantially underestimate their total social value. By providing the first individual–level causal estimates of the educational effects of a congestion policy, this study strengthens the empirical findings on the broader role of air pollution on children's development.

Keywords: Congestion pricing; Air pollution; Student achievement; Educational inequality

The Effect of School Green Policies on Student Outcomes: A Counterfactual Impact Evaluation Using ICCS 2022 Data

Lorenzo Cerullo Bruno – Daniele Bondonio – Daniele Vidoni

Introduction

Europe has set itself an ambitious environmental agenda. The European Green Deal commits the Union to climate neutrality by 2050, and both the Green Deal and the European Climate Pact make clear that regulation alone will not be enough: the transition also depends on changes in everyday values, choices and behaviour (European Commission, 2019, 2020). Survey evidence shows that Europeans are aware and concerned: 78% say environmental problems affect their daily life and health (European Commission, 2024), yet this awareness does not translate easily into action. Among young Europeans, more than four in five think changing their own lifestyle for the environment matters, but fewer than a third report actively doing so (European Commission, 2024). Closing this gap between concern and behaviour has become a central policy concern.

Schools reach almost every young person, repeatedly, during the years when values and habits form. The 2022 Council Recommendation on learning for the green transition, the GreenComp competence framework and the 2024 Education and Training Monitor all position schools as active players in the green transition, where sustainability should be lived through everyday institutional practice in a "whole–school approach" and not only taught as classroom content (Council of the European Union, 2022; Bianchi et al., 2022; European Commission, 2024). The ambition, however, is not yet matched by empirical data: schools vary widely in how far they adopt a green profile, and only a minority of young people feel they have a good chance to learn about sustainability at school (European Commission, 2024).

Four strands of research frame the question this study addresses. The first treats environmental dispositions as genuine outcomes of schooling, much as civic and citizenship research treats democratic attitudes, with schools shaping not only what students know but what they value and do

(Isac et al., 2014; Ambasz et al., 2025). The second stresses that "being green" is not a single thing: studies using recent international data separate environmental concern, pro–environmental attitudes, willingness to act and reported behaviour, and show that these do not always move together (Wasilewska, 2023; Isac et al., 2025; Braun et al., 2018), which is why this study treats attitudes, intended activism and behaviour separately. The third strand looks inside the school: open classroom discussion, hands–on teaching, eco–clubs and concrete environmental policies have all been linked to greener student outcomes (Penrod et al., 2025; Boeve–de Pauw & Van Petegem, 2018; Mónus, 2022), although this work is largely correlational and struggles to separate a genuine school effect from the family background of the pupils a school happens to enrol (Mónus, 2022). A fourth strand moves from associations to more direct tests of whether environmental education is effective: a meta–analysis confirms that such programmes shift knowledge, attitudes and behaviour, though effects vary widely and lean heavily on self–reports (van de Wetering et al., 2022); a longitudinal study ties students' action competence to sustained sustainability teaching (Olsson et al., 2022); and a randomised experiment in Chile finds sizeable gains in children's knowledge, attitudes and practices (Jaime et al., 2023). Evidence closer to whole–school profiles is scarcer but consistent, with green–school certification (Salazar et al., 2024) and environmental clubs (Houmam & Aomar, 2023) both linked to greener behaviour. Even so, these stronger designs cannot fully escape the validity limits of educational inference drawn from observational data (Rutkowski & Delandshere, 2016).

This study addresses that gap with recent cross–national data and a design built to isolate the school–level effect: it asks whether the school itself makes a difference once pre–existing student characteristics are held constant, or whether green schools simply enrol students who were already greener to begin with.

Data used

The study draws on ICCS 2022, the IEA International Civic and Citizenship Education Study, restricted to 21 European education systems and 68,391 lower–secondary students (typically Grade 8). The 2022 cycle added sustainability as a formal subdomain of its civic knowledge test and made it one of its focus areas (Schulz et al., 2023), which suits this question well. The definition of a "green school" is built from what principals report about their own school (School Questionnaire, Question 10: "To what extent are the following practices implemented at this school?"), which mitigates common–method bias, where students would otherwise report both the cause and the effect. Four practices that students can actually see in daily school life were selected: recycling bins, energy–saving measures, activities that encourage environmentally friendly behaviour, and a school garden. A school counts as green when its principal reports the strongest level of implementation on at least three of these four, the aim being to capture a consistent environmental ethos rather than one single isolated initiative.

The outcomes are chosen to respect the multidimensional character of greenness. On the non–cognitive side, three continuous scales are used (attitudes toward environmental protection, expected future activism and self–reported sustainable behaviour) together with four categorical variables capturing specific environmental practices and attitudes (participation in activities to protect the environment, importance of an eco–friendly lifestyle, and how often students choose or refuse particular products). On the cognitive side, the study uses students' general civic–knowledge score (five plausible values) and a dedicated sustainability–knowledge (ESD) score. Including the ESD score is a key contribution: it applies a rarely used scale that tests whether green schools foster environmental literacy or just awareness.

Because greener schools tend to enrol different students, the analysis controls for the factors the literature flags as the main confounders. At student level, the model holds constant gender, age, socio–economic status and immigrant background, since individual background is one of the most consistent predictors of both civic proficiency and environmental dispositions: girls and pupils from more advantaged families tend to hold stronger environmental values and a greater willingness to act (Franzen & Vogl, 2013; Gifford & Nilsson, 2014; Isac et al., 2022; Wasilewska, 2023), while

students with an immigrant background may face barriers linked to lower civic knowledge and weaker attitudes (Isac et al., 2014; Isac et al., 2025). At school and classroom level, the analysis accounts for school location, social composition and the openness of classroom discussion, since urban and rural students differ in both their environmental concern and their reported behaviour (Gifford & Nilsson, 2014; Mónus, 2022), a school's social composition is itself a known driver of pupils' consumption choices and pro–environmental behaviour (Mónus, 2022), and an open climate supports both civic learning and readiness for climate action (Campbell et al., 2024; Penrod et al., 2025).

Methodology

The estimates come from a counterfactual impact evaluation based on propensity score matching (Rosenbaum & Rubin, 1983), where the propensity score is the conditional probability of receiving treatment given a set of observed covariates. In plain terms, each student in a green school is paired with a closely comparable student in a non–green school, matched on the characteristics listed above, so that the remaining difference in outcomes can be attributed to the school's green profile rather than to who the students are. Matching used a nearest–neighbour algorithm (1:1, with replacement) with a caliper of 0.01; alternative calipers left the results unchanged.

Since the matched pairs are compared within each country, national differences in curricula and policy do not drive the results: the propensity model was estimated separately for each education system, and the country–specific effects were then combined into a single weighted "Global ATT" (average treatment effect on the treated). Civic knowledge is measured with the five plausible values supplied by ICCS 2022, used jointly so that measurement error is properly reflected (Wu, 2005).

Robustness checks (a pooled specification with country fixed effects, alternative "green school" definitions and alternative splits of the categorical outcomes) left the main pattern unchanged. Some limitations remain: ICCS 2022 is cross–sectional, so it cannot track students over time; attitudes and behaviour are self–reported; the "green school" measure is a constructed threshold; and matching cannot rule out unobserved differences such as teacher quality or parental motivation. The estimates are therefore more credible than a simple comparison of averages, but fall short of the certainty of a true experiment.

Results and discussion

The analytical approach differs from a simple comparison of all students: it matches individual green–school students with observationally equivalent peers in non–green schools, as described above. The estimates are summarised below.

Outcome	Effect	Meaning
Environmental attitudes	**+0.40 pts *****	Greener attitudes.
Environmental activism	**+0.33 pts ****	Slightly more inclined to act in future.
Sustainable behaviour	**+0.32 pts ****	Slightly greener everyday habits.
Eco–friendly lifestyle	**+1.9 pp ****	About 2 in 100 more students rate it the top priority.
Protect environment, Refuse harmful products, Buy sustainable products	**+0.2 – +0.8 pp (n.s.)**	No reliable effect on these three categorical variables.
Civic knowledge test (5 plausible values)	**+1.6 – +2.4 pts (n.s.; PV3 *)**	No reliable effect on what students know.
Sustainability knowledge: ESD (test score)	**+2.4 pts ***	No reliable effect on sustainability knowledge.

*Notes: continuous attitude and behaviour scales have a mean of 50 and a standard deviation of 10; the knowledge tests have a mean of 500 and a standard deviation of 100. "pts" = points on standardised scales; "pp" = percentage points. Significance: *** p < 0.01; ** p < 0.05; * p < 0.10; n.s. = not statistically significant. Effects are the average difference between students in green schools and matched comparison students, pooled across 21 education systems.*

Two patterns stand out, and both were expected. Across the three attitude–and–behaviour scales, students in green schools come out ahead of their matched peers by a small but statistically reliable margin. Each gap on its own is modest; what strengthens the picture is that all three point the same way, and that the largest applies to environmental attitudes. Among the more specific yes/no measures, only one moves: students in green schools are about two percentage points more likely to say that changing your lifestyle for the environment is very important, with the other specific behaviours pointing in the same direction. On the cognitive side there is no detectable effect. Neither general civic knowledge nor the sustainability–specific score moves in a statistically meaningful way; the estimates are positive but small and do not clear the usual significance bar. The reading is straightforward: the everyday, organisational features of a green school are enough to nudge attitudes and behaviour, but not, on their own, to raise measured knowledge. This is in line with earlier work showing that action competence develops through active teaching (curriculum integration, sustainability lessons, participatory learning) rather than the passive school environment (Olsson et al., 2022; Boeve–de Pauw & Van Petegem, 2018).

Two specific results deserve a note of caution. The estimates for activism and for rating a green lifestyle as important both lose significance when the model is specified differently, so they are best read as suggestive. The core finding (small positive effects on attitudes and behaviour, no effect on knowledge) holds up across the alternative checks.

These results point to a clear conclusion for policymakers. Family background and social composition weigh heavily on students' environmental outcomes, but they are not levers a ministry can act on, while visible green practices are. What the evidence does not support is the expectation that such practices will, by themselves, raise what students know about sustainability: if the goal is measurable learning gains, organisational interventions are not a substitute for teaching. Visible green practices instead look like a low–cost, low–risk way to nudge the attitudes and behaviour that Europe's climate

strategy is trying to shift, and a complement to instruction–focused education for sustainable development. One caveat applies throughout: the overall estimate is an average across very different national systems, and the country–by–country effects vary considerably in size and even in direction; the general European tendency is positive, but how well low–cost green practices work in any single system likely depends on local institutional and cultural conditions.

Conclusions

Across 21 European education systems, secondary students who attend green schools hold modestly but consistently greener attitudes and report greener behaviour than comparable students elsewhere, once individual, classroom and school background are taken into account. The effect is small, specific to attitudes and behaviour, robust to alternative specifications, and obtained at very low cost, since it does not come at the expense of measured knowledge. For education systems looking for feasible ways to support the green transition through schools, this combination (cheap, simple to introduce, no academic downside, with a real if modest push) makes everyday green practices a credible starting point, provided they are integrated with serious teaching for sustainable development.

Keywords: Green schools; Environmental attitudes; Propensity score matching; ICCS 2022

Indirect effects of grade retention on student achievement

Dalit Contini – Guido Salza – Benedetta Valpreda

Introduction

Grade repetition is a measure used in many countries to deal with situations of poor academic performance by some students. Theoretically conceived as an opportunity to address educational deficits before proceeding further in the academic pathway, repetition is often perceived (and experienced) by students as a demonstration of incompetence or even personal "failure".

Scientific research has often debated the usefulness of retention as a recovery tool. Most of the existing literature focuses on the effect on students at risk of retention. In countries where this measure is allowed in primary schools, some studies show that retention has positive effects on the academic performance of retained students, although these effects often diminish after the first years [1, 2, 3]. In high school, however, the prevailing evidence is that retention has negative effects on students' careers and well–being. For retained students, dropout rates increase [4, 5], the likelihood of enrolling in university decreases [6, 7], and self–esteem, motivation, and self–concept worsen [8].

While the effects of retention on retained students are often debated, evidence on its effects on the general student population is even more limited. Indeed, the very existence of retention policies can influence all students. Retention can act as an implicit warning: those who do not achieve sufficient results risk repeating the academic year. If students perceive this situation as a challenge to overcome, the perceived "threat" of retention may motivate them to engage more in their studies. Conversely, if achieving the required results seems unattainable, the threat of retention can have the opposite effect, leading to anxiety, stress, and emotional strain, especially among the most sensitive students. Moreover, grade retention may prompt retained students to transfer to another school, making the classroom more homogeneous and potentially improving the overall learning atmosphere.

Empirical evidence on the indirect effects of retention is limited and inconclusive. Koppensteiner [9] finds that automatic promotion in Brazil lowers grade 4 performance, particularly for at–risk students, suggesting it disincentivizes effort. A Belgian study [10] shows mixed results, with improvements in some cases but little impact on vulnerable students. Jacob and Lefgren [11] find that a policy linking promotion to test scores benefits third graders briefly but has no lasting effect,

while sixth graders see no impact. In Italy, Battistin and Schizzerotto [12] report performance declines in technical schools after a 2008 reform reintroducing retention after remedial exams in the fall, but no effect in academic high schools.

Following this line of inquiry, we ask: what is the indirect effect of grade retention on the achievement of the general student population, and in particular of those with satisfactory or high academic performance? Our article contributes to this literature by estimating these effects in the Italian context.

Context, Data and Analytical Strategy

In Italy, grade repetition is rare in primary and middle school but rises sharply in high school; in 2018–19, nearly 12% of ninth graders were retained. Teachers assign final subject grades annually, and to advance students need at least 6/10 in all subjects, including behavior. If grades are low, teachers assess the student's potential to meet expectations, allowing for considerable discretion in promotion decisions. Assuming that the propensity to retain varies at the school level, following Contini and Salza [5] and Valpreda [13], we estimate the indirect effect of retention on learning by leveraging differences in school "strictness", understood as the propensity of schools to retain students, after controlling for student and school characteristics.

This study uses a rich longitudinal dataset on students who entered grade 9 in 2014/15 in Lombardy, Piedmont, and Veneto, covering 2013/14 to 2016/17. Data come from the National Student Register (ANS) and the National Institute for the Evaluation of the Schooling System (INVALSI), including school and student details, end–of–year grades in Italian, mathematics, and behavior, standardized INVALSI test scores (grades 8 and 10), and family background. The analysis focuses on students promoted to grade 10 who took the INVALSI test in spring 2016.

The ideal setting for evaluating the indirect effects of retention is the introduction of a policy that either implements or restricts its use. This allows for a comparison of similar students in different contexts: one where retention is possible and one where it is not. Estimating a causal effect is more complex when there are no policy changes, or retention is assigned without a transparent rule (e.g., based on standardized test results). Since the decision to retain in Italy is discretionary, schools may differ in their strictness. This variability can be exploited by analyzing the impact of school strictness on students' achievement in otherwise (ideally) identical schools. If, among promoted students, learning is better in stricter schools, retention can be seen as a measure that helps preserve the learning levels of the students in schools.

However, strictness is not observed and must be estimated. The starting point of our estimation strategy is a multilevel logistic model for the probability of being retained in grade 9, with a random component at the school level. Within each school, the decision to promote or retain largely depends on the array of student's performance measures, which, given the richness of the information in our dataset, captures a wide range of cognitive and non–cognitive abilities.

Let T_{ij}^{*} be the latent variable underlying the retention probability, X_{ij} are individual characteristics including gender, school grades (Italian, math, behavior) and Invalsi test scores (Italian, math) as well as parental education and migrant origin; Z_j is a vector of high school characteristics including school curricula, whether public or private, average past grades and standardized test scores, socioeconomic composition and share of females of the student body; $\varepsilon_{ij}$is the individual error term. The retention probability also depends on a school–level random component u_j, conceptualized as the propensity to retain (strictness), net of the observable characteristics of students and school:

$$T_{ij}^{*} = \beta_0 + \beta_1 X_{ij} + \beta_2 Z_j + u_j + \varepsilon_{ij} \quad (1)$$

$$T = \begin{cases} 1 & if\ T^{*} > 0 \\ 0 & if\ T^{*} < 0 \end{cases}$$

Once the strictness of each school $Strict_j = u_j$ has been estimated via Empirical Bayes posterior estimation, we move to the main model, with dependent variable Y_{ij}, the standardized test score in grade 10 (i.e. in the following school year), and (estimated) strictness as the main explanatory variable:

$$Y_{ij} = \gamma_0 + \theta Strict_j + \gamma_1 X_{ij} + \gamma_2 Z_j + v_{ij} \qquad (2)$$

The estimate of θ, the key parameter of interest capturing the causal effect of school strictness on student achievement, will be unbiased if strictness is exogenous. This assumption holds when initial high school choices are based on observable school characteristics included in the model, rather than on the unobserved component of strictness. To address potential endogeneity concerns and measurement error, related to the fact that true strictness is not observed and needs to be estimated, we conduct an extensive simulation study to evaluate these estimation challenges (Section 4).

Results

To allow meaningful comparisons, both steps (1) and (2) are estimated separately by school track. Table 1 presents the main results. In general, the strictness of the school does not appear to influence achievement in grade 10. However, a positive effect is observed in non–traditional lyceums: controlling for a wide array of individual and school characteristics, as the strictness of these schools rises, student performance improves. We also limited the analysis to students with a low retention probability (less than 5%) even in schools with strictness above the 95° percentile (right panel of Table 1). This prevents the inclusion of underperforming students in more lenient schools, who would not be well–matched to students in stricter schools due to different latent ability thresholds for promotion. In other lyceums we still find a positive effect but of slightly lower size.

Table 1 – Effect of strictness on students' test scores in grade 10

	All promoted students		Students at low risk of retention	
	Italian	**Mathematics**	**Italian**	**Mathematics**
Traditional lyceums	0.013	0.077	–0.012	0.057
Other lyceums	0.186 ***	0.172 **	0.155 ***	0.144 **
Technical track	–0.022	0.006	–0.017	–0.009
Vocational track	0.015	–0.023	0.041	0.090

NOTES. *p<0.10; **p<0.05; ***p < 0.01. Standard errors cluster by school.

Limitations and simulation study

As mentioned above, we must address two problematic issues.

(a) In the main analyses, we make the simplifying assumption that strictness is exogenous for both models (1) and (2), i.e. $\rho(u_j, \varepsilon_{ij}) = \rho(u_j, v_{ij}) = 0$. We believe this assumption is reasonable because decision–making is based on observable characteristics, whereas strictness (as defined here) is inherently unobserved. Moreover, we can account for a rich array of students and school characteristics. However, if students possess some knowledge of strictness, it could introduce endogeneity. In this case, $\rho(u_j, -\varepsilon_{ij})$ and $\rho(u_j, v_{ij})$ are likely positive, because, if a relationship exists, motivated students are more likely to choose stricter schools. In this scenario, we should end up overestimating θ.

A potential solution is to use an IV approach, with a viable instrument being the school's share of students who transferred to a different school after grade 9, based on the previous birth cohort's experience. This share correlates with (estimated) strictness and can be assumed not to influence the test scores of a different cohort. However, it proves to be a weak instrument (F–statistic = 4.5, well

below the commonly used threshold of 10 for instrument validity). Moreover, its use may be challenged due to the issue described next [13].

(b) Measurement error in the treatment variable in (2), arising from the fact that strictness is unobserved and must be estimated, may introduce bias in the estimate of θ even in the exogenous scenario. In the case of classical measurement error, this would lead to an underestimation of the effect of strictness.

To investigate the direction and magnitude of these biases, we conducted a series of simulations based on a simplified version of models (1) and (2), where X_{ij} is a scalar variable capturing ability (standard normal), and there are no observed school–level variables. Following (1), we assign grade retention to the 15% of students (close to the observed share in the population) with the highest probability of being retained across all schools.

For each school, strictness is set as $u_j = mw$, where $m > 0$ varies across scenarios and w is standard normal. For each student, unobserved motivation, ε_{ij} in (1) and v_{ij} in (2), are set respectively to $e_1(N(0,1) + kw)$ and $e_2(N(0,1) + kw)$, where $e_1 < 0$ and $e_2 > 0$.[2] This specification also allows to manipulate endogeneity: $\rho(u_j, \varepsilon_{ij}) = \rho(u_j, v_{ij}) = 0$ when $k = 0$, and $\rho(u_j, -\varepsilon_{ij}) = \rho(u_j, v_{ij}) = \frac{k}{\sqrt{1+k^2}}$ when $k > 0$. The implication is that, with $k > 0$, more motivated students would self–select into stricter schools. The comparison of interest is between the (standardized) true strictness effect $\theta \geq 0$ and its estimate $\hat{\theta}$. Key results are qualitatively summarized in Table 2.

Table 2 – Biased estimates under different scenarios

Strictness is…	*Known*	*To be estimated*
Exogenous	Unbiased estimates	Downward bias
Endogenous	Upward bias	Mixed results

In greater detail, we review the results when strictness is unknown and must be estimated from equation (1). When strictness is exogeneous ($k = 0$), its effects is always underestimated ($\hat{\theta} < \theta$). However, the bias is sizable only under specific conditions – namely, when e_1 is large (the unobserved individual component plays a substantial role in retention), m is small (implying low variability in strictness, thus, having little impact on the probability of grade retention) and θ is large, meaning that strictness strongly affects the performance of promoted students. Nonetheless, we always reject the null hypothesis of no effect of strictness when $\theta \neq 0$, and we never reject it when $\theta = 0$. That said, we believe this scenario is unrealistic, as it represents a case where differences in strictness are minimal, yet school strictness has a substantial impact on performance.

Preliminary results from endogenous scenarios reveal mixed outcomes. The true effect of strictness is sometimes underestimated and, in other instances, overestimated. Notably, strong underestimation occurs when e_1 is large and m is small, as in the exogenous case, and k is very large (as mentioned above, we believe this scenario is unrealistic). In this case, due to strong endogeneity in (1), strictness is poorly estimated and the prevailing bias arises from measurement error. Conversely, strong overestimation occurs when e_1 is small, m is large and k is large. Here, the dominant bias stems from strong endogeneity in (3).

Conclusions

In conclusion, the analyses suggest that grade retention not only negatively impacts retained students [5], but also, in most cases, fails to benefit promoted students in terms of future academic

[2] We set $e_1 < 0$ to reduce the likelihood of grade repetition for more motivated students and $e_2 > 0$ to reflect their better performance.

performance. The only exception is observed in non–traditional lyceums, where there is evidence of significant positive effects.
The simulation study, designed to assess potential biases from measurement error in the strictness variable and violations of its exogeneity assumption, indicates that: (i) the effect of school strictness is unlikely to be largely underestimated; and (ii) its effects may, in fact, be substantially overestimated. These findings reinforce the conclusion that grade retention is not an effective policy for addressing student underperformance.

Keywords: Grade retention; Test scores; Causal effects

Teachers' Digital Readiness and Student Outcomes: Cross–National Evidence from TALIS and ICILS

Letizia Gambi – Daniele Vidoni

Background and Policy Relevance
Teacher self–efficacy (TSE) is considered an important determinant of instructional quality, classroom practices, and student learning outcomes. A large body of literature shows that teachers' beliefs in their own teaching capabilities are associated with student engagement, motivation, and academic achievement (Zee & Koomen, 2016). However, more recent evidence suggests that this relationship is not always straightforward and may depend on how self–efficacy is conceptualised and measured, particularly when distinguishing between general and domain–specific forms (Bardach & Klassen, 2021).
The paper builds on several strands of literature. First, research on teacher self–efficacy highlights its importance for instructional quality and student outcomes (Zee & Koomen, 2016), although recent meta–analyses suggest that the relationship may be weaker and more context–dependent than previously assumed (Kang et al., 2025). Bardach and Klassen (2021) further emphasise that domain specificity is crucial for understanding the effects of teacher self–efficacy. Second, research on digital competence and professional development highlights the importance of ICT–focused training and school support structures in shaping teachers' confidence and instructional practices (Rich et al., 2021; Kruskopf et al., 2024; Reeves & Hamilton, 2022). Third, recent comparative evidence from large–scale assessments indicates that the relationship between teacher self–efficacy and student achievement is often modest and sensitive to model specification and measurement choices (Jerrim et al., 2023). This underscores the importance of rigorous identification strategies and cross–dataset validation.
Finally, digital competences have gained particular policy relevance for both students and teachers in the context of recent European Commission initiatives, including the Union of Skills, the Digital Education Action Plan 2021–2027 and the Digital Decade policy programme 2030, which emphasise the role of education in strengthening EU competitiveness and recognise digital competence as one of the key competences for lifelong learning, as defined in the DigComp framework. Furthermore, the European Commission's Digital Education Action Plan emphasises the importance of "digitally competent and confident teachers" as a cornerstone of a high–performing digital education ecosystem (European Commission / EACEA / Eurydice, 2023). Despite this strong policy emphasis, empirical evidence on what drives teachers' digital self–efficacy remains limited, especially in comparative European settings. Even less is known about whether and how digital self–efficacy translates into improved student outcomes with regards to digital competences.
This project addresses these gaps by providing a cross–national comparative analysis of the predictors and consequences of teachers' digital self–efficacy and readiness – understood as teachers' confidence and perceived readiness to effectively integrate digital technologies into

teaching and learning –, focusing on both school and teacher–level predictors, and student–level outcomes across Europe.

Research Objectives

The project pursues two main research objectives. First, it investigates the determinants of teachers' digital self–efficacy as a proxy for digital readiness. The analysis examines how teachers' academic background, professional qualifications, and participation in pre–service and in–service training influence their self–efficacy beliefs. A particular distinction is made between general training and ICT–specific professional development, as well as subject–specific preparation. In addition, the study considers the role of school–level conditions, including leadership quality, availability of digital infrastructure, and broader school climate. These factors are expected to shape not only general teacher self–efficacy but also domain–specific forms such as digital readiness.

Second, the paper investigates the relationship between teachers' digital self–efficacy and students' ICT competences. It assesses whether teachers who report higher levels of digital confidence are associated with higher levels of student performance in computer and information literacy. The analysis further explores heterogeneity in this relationship across teacher characteristics (such as experience, gender, and STEM versus non–STEM background) and school contexts.

Data Sources, Comparative Design and Conceptual Framework

The empirical strategy is based on the integration of two international large–scale assessments: the Teaching and Learning International Survey (TALIS, teacher and school leaders' questionnaire, ISCED 2) and the International Computer and Information Literacy Study (ICILS, grade 8 students and teachers). These datasets provide complementary perspectives on teachers, students, and school environments across European education systems. This multi–survey design allows for a uniquely rich comparative framework, enabling the analysis of teacher self–efficacy across different subject domains and institutional contexts while linking teacher–level measures to student outcomes in a consistent cross–national framework.

The conceptual framework distinguishes between general teacher self–efficacy and digital self–efficacy, understood as teachers' perceived readiness to effectively integrate digital technologies into teaching practice. It further incorporates measures of teacher motivation, school climate, and instructional context, recognising that self–efficacy is shaped by both individual and institutional factors.

A key methodological challenge is ensuring comparability of constructs across different international assessments. To address this, the study develops a harmonised measurement framework that aligns comparable indicators across TALIS and ICILS. Where necessary, latent constructs are derived using exploratory and confirmatory factor analysis, ensuring internal consistency and conceptual validity across datasets. Existing OECD–provided scales, such as TALIS teacher self–efficacy measures, are combined with newly constructed indices to ensure comparability. All constructed variables are standardised to mean zero and unit variance. Standardisation is implemented within countries for national analyses and across countries for pooled models, ensuring both comparability and interpretability of estimates (Jerrim et al., 2025).

Empirical Strategy

Given the correlational nature of TALIS 2024 survey data, the paper adopts a selection–on–observables framework to strengthen causal interpretation of the empirical analysis. This includes multilevel models[3] with extensive controls, complemented by propensity score matching (and robustness checks using alternative weighting schemes) to assess the sensitivity of results to observable selection bias.

[3] Multilevel models account for the hierarchical structure of the data, with students nested within schools and countries, and estimate the association between teacher self-efficacy—particularly digital self-efficacy—and student outcomes, while controlling for relevant individual- and school-level covariates.

In addition to these identification strategies, we incorporate multilevel structural equation modelling to examine potential mediation pathways between teacher training, self–efficacy, and student outcomes. Machine learning methods such as random forests are also used to identify the most important predictors of teacher self–efficacy and student achievement in a data–driven framework. This combination of theory–driven and exploratory approaches allows for a robust and comprehensive understanding of the relationships under study.

Expected Contributions and Hypotheses

This paper contributes to three main areas of research. First, it provides a comprehensive cross–national analysis of the predictors of teacher digital self–efficacy, as well as of the relationship between teacher self–efficacy and students' digital competences. Second, it demonstrates the value of using multiple international large–scale assessments to generate robust and comparable evidence across different measurement frameworks. By integrating these data sources, the study provides a novel empirical basis for understanding how teacher–level beliefs relate to student outcomes across diverse educational contexts, thereby contributing to ongoing policy debates on educational quality and equity in an increasingly digital learning environment.

Based on existing literature, we hypothesise that ICT–focused professional development and school–level digital policy and leadership support are key predictors of digital self–efficacy. It further expects that the relationship between teacher self–efficacy and student outcomes will be positive but modest in magnitude, and stronger for domain–specific forms of self–efficacy – such as digital self–efficacy – than for general measures.

Keywords: Basic skills; Digital competences; Self–efficacy; TALIS

Can Teacher Training in Active Learning Reduce the Math Gender Gap? Evidence from a Randomized Controlled Trial in Italy

Maria Laura Di Tommaso – Daniela Piazzalunga – Giulia Savio

Introduction

Despite substantial progress in gender equality in education, mathematics remains one of the few domains where girls continue to underperform relative to boys in most developed countries. International assessments consistently document a gender gap in mathematics achievement (OECD, 2023). The gap is particularly pronounced among high–achieving students (Ellison and Swanson, 2010, 2023; Fryer and Levitt, 2010), contributing to women's lower representation in STEM fields and subsequent labor–market disparities (Turner and Bowen, 1999; Delaney and Devereux, 2019; Card and Payne, 2021; Francesconi and Parey, 2018).

A broad literature attributes these differences to parental and teacher stereotypes (Alan et al., 2018; Carlana, 2019; Dossi et al., 2021), gender differences in confidence and mathematics anxiety (Ho et al., 2000; OECD, 2015; Di Tommaso et al., 2021), competitive preferences (Gneezy et al., 2003; Niederle and Vesterlund, 2010), and exposure to female role models (Dee, 2007; Coenen et al., 2018). Another potentially important mechanism concerns classroom pedagogy.

Our theoretical framework builds on constructivist theories of learning, according to which knowledge is actively constructed through exploration, collaboration and discussion rather than passively transmitted by teachers (Dweck, 2006; Boaler, 2013; Gutierrez and Boero, 2006). Active learning environments are expected to foster engagement, reduce anxiety, improve confidence and promote deeper conceptual understanding. These mechanisms may be especially beneficial for girls, who often report lower confidence in mathematics despite comparable ability.

Although educational research has documented the benefits of learner–centred approaches (Slavin and Lake, 2008), rigorous causal evidence from economics remains limited. Existing studies show

that teaching practices matter for cognitive achievement (Schwerdt and Wuppermann, 2011; Bietenbeck, 2014; Lavy, 2016), but little is known about whether teacher training in active learning can reduce gender gaps in mathematics. Moreover, most previous interventions have been implemented directly by researchers rather than classroom teachers.
This paper evaluates a teacher training program promoting active–learning mathematics instruction among Grade 3 primary–school teachers in Italy. Using a stratified cluster randomized controlled trial conducted in Piedmont and Campania, we estimate the causal effect of teacher training on teachers' practices and students' mathematics achievement. The paper contributes by evaluating a scalable intervention delivered through teachers and by examining heterogeneous effects by gender and baseline achievement. We find that the intervention was effective in rising math scores only among female students.

Data

The analysis uses data from a stratified cluster randomized controlled trial implemented in Piedmont and Campania. Grade 3 mathematics teachers volunteered to participate and school–buildings were randomly assigned to treatment or control. The intervention consisted of nine sessions introducing active–learning methodologies based on the Italian Laboratorio di Matematica approach. The study includes 76 teachers, 81 classes and 1,396 pupils; the final teacher sample contains 61 teachers, and 1,120 pupils because the combination of attrition and the exclusion restriction of the sample. Teacher surveys measure pedagogical beliefs and classroom practices, while student questionnaires include mathematics tests and non–cognitive measures.

Methodology

Treatment assignment occurred at the school–building level to minimize spillover effects. Treatment effects are estimated using OLS regressions controlling for baseline cognitive ability (or teacher ability), background characteristics, and strata fixed effects. The equation for teachers is as follows:

$$Y_{jt} = \alpha + \beta T_j + \theta Y_{j,t-1} + X_j'\gamma + \mu_s + \epsilon_{jt}$$

and the equation for students is as follows:

$$Y_{it} = \alpha + \beta T_i + \delta(T_i \times Y_{i,t-1}) + \theta Y_{i,t-1} + X_j'\gamma + \mu_s + \epsilon_{it}$$

Our main outcome variable in the teacher equation is the teaching practices scale, which combines measures of teacher–centered and student–centered (reverse–coded) teaching practices. The scale is standardized, with higher values indicating more teacher–centered teaching. In the student equation, the main outcome variable is the cognitive index, calculated as the sum of correct answers to 37 cognitive questions (1 = correct, 0 = incorrect) and standardized to have mean zero and unit variance. The set of covariates in both equations includes socio–demographic characteristics such as gender, nationality, and experience. In the student equation, we additionally control for classroom characteristics. The outcome variable in the students' equation is the performance in the math test, where test scores are standardized.

We estimate intention–to–treat (ITT) effects. Standard errors are clustered at the school–building level. We also examine heterogeneous effects by estimating the baseline specification separately for subsamples defined by gender.

Results and Discussion

We report the effects of the intervention on teachers' instructional practices and on students' mathematics achievement. Results first present average treatment effects and then heterogeneous effects by gender.

Regarding the impact on teacher practices, Table 1 shows that after the intervention the teacher–centered practices were reduced between 0.42 and 0.55 SD. These results suggest that the intervention substantially reduced teacher–centered practices.

Table 1: Teachers' practices

	Teacher-centered practices scale	
	(1)	(2)
Allocation to treatment	-0.423**	-0.545***
	(0.208)	(0.186)
Mean	0.006	-0.072
Observations	55	55
Strata FE	Yes	Yes
Covariates	Pre	Yes

Regarding the impact on students' math performance, Table 2 shows that after the intervention, on average, there was not improvement on mathematics achievement. We also observe that relevant of pre–cognitive ability and gender status predict the current performance of students.

Table 2: Cognitive Ability (Full Sample)

	Standardized Cognitive Test				
	(1)	(2)	(3)	(4)	(5)
Allocation to Treat	-0.031	-0.025	-0.018	-0.056	-0.053
	(0.074)	(0.058)	(0.058)	(0.059)	(0.059)
Std. Outcome Pre		0.862***	0.824***	0.806***	0.762***
		(0.039)	(0.040)	(0.064)	(0.063)
Treat x Std. Outcome Pre				0.105	0.118
				(0.077)	(0.076)
Female - Pre			-0.132***		-0.138***
			(0.048)		(0.047)
Mean	-0.040	-0.040	-0.040	-0.040	-0.040
Observations	1120	1120	1120	1120	1120
Strata FE	Yes	Yes	Yes	Yes	Yes
Covariates	Sh.F.	Sh. + Pre	Yes	Sh. + Pre	Yes

Given the previous results, we explore potential heterogeneous effects by gender condition. Tables 3 and 4 are the baseline regression for a subsample of girls and boys, respectively. We find that after the intervention the math performance for the subsample of girls improve conditional on their

previous cognitive ability by 0.24 SD, whereas we do not find any change for boys. These results suggest that the impact of the intervention are concentrated within girls.

Table 3: Cognitive Ability (Girls)

	Standardized Cognitive Test				
	(1)	(2)	(3)	(4)	(5)
Allocation to Treat	0.033	-0.025	-0.027	-0.057	-0.062
	(0.082)	(0.058)	(0.059)	(0.059)	(0.061)
Std. Outcome Pre		0.746***	0.733***	0.633***	0.614***
		(0.052)	(0.045)	(0.082)	(0.075)
Treat x Std. Outcome Pre				0.226**	0.236**
				(0.094)	(0.091)
Mean	-0.173	-0.173	-0.173	-0.173	-0.173
Observations	559	559	559	559	559
Strata FE	Yes	Yes	Yes	Yes	Yes
Covariates	Sh.F.	Sh. + Pre	Yes	Sh. + Pre	Yes

Table 4: Cognitive Ability (Boys)

	Standardized Cognitive Test				
	(1)	(2)	(3)	(4)	(5)
Allocation to Treat	-0.103	-0.013	-0.013	-0.011	-0.009
	(0.095)	(0.076)	(0.071)	(0.082)	(0.082)
Std. Outcome Pre		0.956***	0.928***	0.970***	0.942***
		(0.052)	(0.055)	(0.099)	(0.099)
Treat x Std. Outcome Pre				-0.025	-0.025
				(0.110)	(0.112)
Mean	0.094	0.094	0.094	0.094	0.094
Observations	561	561	561	561	561
Strata FE	Yes	Yes	Yes	Yes	Yes
Covariates	Sh.F.	Sh. + Pre	Yes	Sh. + Pre	Yes

Conclusions

This study evaluates whether teacher training in active learning can improve mathematics learning and reduce gender disparities in primary school. By focusing on a scalable intervention delivered through regular teachers, the paper contributes to the literature on teacher effectiveness,

mathematics education and gender gaps in educational achievement. The results will provide evidence relevant for the design of teacher professional development policies.

Keywords: Grade retention; Test scores; Causal effects

Evaluating Teacher Professional Development to Improve Spatial Reasoning and Mathematics: Evidence from the Matabì Cluster Randomized Trial

Barbara Romano – Adriana Di Liberto – Ludovica Giua – Giovanni Piumatti

Policy problem: mathematics fragility and the early gender gap

Mathematics competence is one of the strongest predictors of later educational attainment, tertiary completion and access to qualified employment, and it underpins the cognitive and metacognitive skills — working memory, planning, self–regulation — that support learning across domains. In Italy, large–scale assessments point to a structural weakness in this area: recent INVALSI results show a persistent decline in average mathematics scores and a growing share of pupils who fail to reach the basic proficiency level by the end of primary school, with territorial and socio–economic disparities widening in later grades. OECD PISA 2022 data confirm the pattern internationally: Italian 15–year–olds score below the OECD average in mathematics, and the gender gap in favour of boys — over 20 points, roughly six months of learning — is among the largest in Europe.

This is a twofold policy problem. On one hand, Italian pupils show an overall weakness in mathematical preparation; on the other, a gender gap opens up in the earliest years of primary school and tends to widen over time, rather than narrowing. The 2025 INVALSI report shows that already in primary school a significant share of girls fail to reach basic proficiency, while the share of high–achieving pupils — and especially girls — in the top proficiency bands is shrinking. Left unaddressed, this early divergence translates into lower female participation in STEM pathways and, over the long run, a loss of national human capital and competitiveness. Closing the gap early is therefore not only a matter of educational equity, but a lever for human–capital development and for the effective, efficient use of public investment in education.

2. The Matabì program: a teacher–mediated, playful–learning intervention

Matabì is a professional–development and classroom intervention, developed by Fondazione Agnelli, that uses LEGO® DUPLO bricks as a pedagogical tool to strengthen **visuo–spatial skills** — an empirically established precursor and correlate of mathematical competence — while embedding explicit spatial language and playful, manipulative learning into everyday teaching practice. Rather than delivering materials directly to pupils, the program is designed as a **train–the–trainer** model: teachers attend five training modules (10 hours total) combining pedagogical content on spatial cognition and playful learning with an explicit module on implicit gender bias in STEM, and subsequently lead structured brick–based activities and three expert–designed workshops in their classrooms. The underlying theory of change is sequential: a genuine transformation of teachers' own spatial skills and spatial language is the necessary first step, which then — and only then — enables measurable gains for pupils, first in visuo–spatial ability and, with a lag, in mathematics, self–efficacy and math anxiety.

3. Evaluation design: a two–wave cluster–randomized controlled trial

Fondazione Agnelli evaluated Matabì through a **Randomized Controlled Trial**, the design offering the strongest identification of causal effects in an education–policy setting. Random assignment took place at **classroom level**, stratified by grade (third/fourth primary), within schools that had volunteered to a public call. This design combines internal validity — differences in outcomes can be attributed to the program rather than pre–existing characteristics — with a controlled, homogeneous comparison within the same school.

Pilot wave (2022–23): 5 schools in Turin, 46 second/third–grade classes, 863 pupils and 48 teachers. Classes were randomized into three arms: in–person teacher training in the first term (T1), online teacher training in the second term (T2), and a business–as–usual control group — allowing a first test of implementation intensity as well as of the program itself.
Scale–up wave (2023–24): 11 schools across Piedmont, Campania and Sicily, 83 classes, 974 pupils and 64 teachers, extending the sample to urban, peri–urban, and socio–economically diverse contexts to test transferability. The design was simplified to a single treatment/control comparison, and a digital platform was introduced for data collection. Crucially, this wave also collected *pre/post* standardized visuo–spatial tests for teachers (the Lappan Test of spatial orientation, LA, and the Purdue Spatial Visualization Test of mental rotation, PU), enabling a direct test of the teacher–mediated causal chain.
Outcomes are measured through validated instruments for pupils' visuo–spatial ability, mathematics competence, self–efficacy and math anxiety, with baseline/endline measurement and classroom–clustered standard errors throughout.

4. Results: a teacher–mediated channel, not a uniform average effect

Teachers. In the 2023–24 wave, teachers in treated classes improved their spatial–orientation scores (LA test) by 1.12 points, significant at the 10% level — about a quarter of a standard deviation at baseline — while gains in mental rotation (PU) were smaller and not significant, consistent with the program's specific focus on spatial orientation and path construction rather than pure mental rotation.
Pupils, pilot wave. Classes trained in person and treated in the first term (T1) showed a significant gain in visuo–spatial skills of +0.23 SD ($p<0.05$); classes trained online and treated later in the year (T2) showed a much smaller, non–significant effect (+0.04 SD). No significant effects emerged on mathematics, self–efficacy or math anxiety in this first year — consistent with the program's sequential theory of change, in which spatial gains precede transfer to mathematics. The contrast between T1 and T2 is itself a policy–relevant finding: it points to *implementation intensity* — in–person training, timing, and duration of exposure — rather than mere assignment to treatment, as the active ingredient.
Pupils, scale–up wave. At the aggregate level, none of the four main outcomes showed a statistically significant treatment effect (visuo–spatial: −0.046 SD; mathematics: +0.061 SD; self–efficacy: +0.007 SD; math anxiety: +0.004 SD) — an average null effect. However, **heterogeneity analysis reveals the mechanism behind this null result**: pupils assigned to teachers who measurably improved their own visuo–spatial skills (pre/post) show gains of +0.20 to +0.25 SD in visuo–spatial ability and +0.10 to +0.15 SD in mathematics, relative to pupils of teachers who did not improve. The effect is driven specifically by teachers' progress on the LA (orientation) test, not the PU (rotation) test — mirroring the content of the training itself. Gender–disaggregated results show differentiated pathways: girls assigned to improved teachers gain almost +0.25 SD in visuo–spatial skills, while boys in the same classes show a faster transfer into mathematics performance (+0.42 SD), suggesting that girls and boys may follow different maturation timelines from spatial to mathematical competence.
Taken together, the two waves show that Matabì's impact is not adequately captured by an intention–to–treat average: the program operates through a **teacher–mediated channel**, and its effectiveness depends on whether — and how deeply — teachers actually transform their own spatial skills and classroom practice. This is a methodologically important lesson for the evaluation of teacher–mediated interventions more broadly: average treatment effects can mask substantial, policy–relevant variation tied to implementation fidelity and professional–development depth.

5. Policy implications

The evidence generated by Matabì supports five main policy indications for the design, delivery and monitoring of similar programs:

- **1. Teacher training only works if it changes practice.** Effects on pupils emerge specifically in classes where teachers demonstrably improved their own spatial skills and systematically

integrated spatial language into teaching. Training programs should therefore be designed and funded around cycles of practice, coaching and structured reflection — not around hours of training delivered.

- **2. Spatial thinking should be recognized as a cross–cutting competence.** Its enhancement is associated with gains not only in spatial tasks but, more modestly, in mathematics, with literature suggesting further spillovers into reading comprehension and logical reasoning. Curricula and teacher–training frameworks should integrate spatial thinking alongside literacy and numeracy as a core competence.
- **3. Gender equity should be an explicit dimension of teaching quality.** Differentiated trajectories for girls and boys, together with the role of self–efficacy, math anxiety and teachers' implicit bias (measured via IAT) in shaping classroom participation, indicate that equity, bias–awareness and gender dynamics should be an explicit, assessed component of teacher professionalism — not an implicit byproduct of good teaching.
- **4. School–family collaboration is essential to equal learning opportunities.** Both waves show that home manipulative play is associated with better outcomes, and that boys use bricks at home more than girls even after in–school exposure. Without targeted family engagement — take–home kits, shared activities, targeted communication — the out–of–school environment risks amplifying gaps that the school is trying to close.
- **5. Monitoring and evaluation systems must integrate outcomes with implementation processes.** Effects depend on implementation quality, intensity of use, coherence with concurrent school projects, and the depth of professional transformation actually achieved — not only on assignment to treatment. Process indicators (regularity of activities, use of spatial language, cooperative practices) are necessary to understand why and under which conditions an intervention works; outcome measurement alone risks masking decisive variation in how schools adopt a given methodology.

6. Ongoing evidence: linking short–term gains to standardized outcomes

A further and methodologically important step in the evaluation is now underway: INVALSI grade–5 standardized assessment data are being collected as each treated cohort completes primary school — school year 2025–26 for the 2022–23 cohort, and 2026–27 for the 2023–24 cohort. Linking Matabì's experimental data to these external, standardized measures — collected two to three years after treatment — will make it possible to test whether the visuo–spatial gains observed at endline translate into more durable effects on mathematics and reading comprehension, offering a more robust assessment of the persistence of impact and directly addressing a common limitation of short–horizon RCTs in education: the risk that early cognitive gains fade, or conversely, compound over time into measurable achievement effects.

7. Contribution to the session

This paper offers a completed two–wave RCT with a clear and policy–actionable identification strategy, an explicit test of a teacher–mediated causal chain rarely observable in education RCTs (via pre/post teacher–level testing), and an ongoing linkage to administrative INVALSI data that will extend the evidence base into a genuine impact–persistence test. It speaks directly to the session's call for rigorous counterfactual evidence combining experimental design with national assessment infrastructures, and for methodological reflection on what average treatment effects can and cannot reveal about how, for whom, and under which implementation conditions basic–skills interventions actually work.

Keywords: RCT; Teacher Trainig; Spatial Abilities; Math

Modelling of the Modelling Process as a Severe Testing Instrument for Causal Inference in Education Policy Evaluation

Giovanni Lari – Daniele Bondonio – Andrea Saltelli

Introduction

Counterfactual impact evaluation has become the reference standard for assessing education policies, but the credibility of any single causal estimate depends on a step that is rarely made explicit: the many analytic choices a researcher makes while building the model that produces that estimate. Sample restrictions, matching algorithms, caliper widths, covariate operationalization are all decisions taken somewhere between data and result, and each of them can move the estimated effect without any single one being individually indefensible. The literature on "researcher degrees of freedom" and the "garden of forking paths" has shown that this multiplicity of admissible choices is a structural feature of applied quantitative research, not an occasional lapse, and that it can generate a wide range of legitimate results from the same data and the same research question.

This contribution proposes Global Sensitivity Analysis (GSA), and specifically the approach known as Modelling of the Modelling Process (MOMP), as a severe testing instrument for counterfactual causal estimates in education policy research. MOMP treats the sequence of modelling decisions itself as a source of uncertainty to be sampled and propagated, alongside the more familiar uncertainty in parameters and data. Framed in Popperian terms, and building on Deborah Mayo's account of severe testing, a causal claim earns credibility not by surviving a single specification but by surviving a systematic exploration of the space of reasonable alternative specifications.

The paper is methodological. It develops the MOMP architecture as a general severe–testing instrument for causal inference, and it uses, as its worked illustration, an ongoing companion study that applies Propensity Score Matching (PSM) to ICILS 2023 data to estimate the effect of heavy versus light social media use on students' computational thinking scores across multiple education systems. The companion study supplies a realistic and non–trivial matching pipeline, with genuine analytic forks (choice of matching algorithm, caliper, replacement, covariate coding, variance estimator), against which the MOMP procedure can be demonstrated. The present abstract reports on the design and methodology of this application; the companion paper is cited as a working reference for readers interested in the substantive education–policy findings.

2. Data used

The illustrative case draws on ICILS 2023 (International Computer and Information Literacy Study), an IEA cross–national assessment that measures computational thinking (CT) and computer and information literacy among lower–secondary students. The companion PSM study defines a heavy–use and a light–use group from students' self–reported social media use, and estimates the average treatment effect on the treated (ATT) on CT scores, following IEA–standard procedures for the combination of plausible values and for cluster–robust inference at school level. The target analytic sample spans the education systems that participated in the CT assessment module.

For the purposes of the present methodological paper, this dataset and pipeline are not an end in themselves but the terrain on which the MOMP exercise is run: what is treated as "data" for the sensitivity analysis is not only the ICILS microdata but the full set of matching specifications that a competent analyst could reasonably have chosen when estimating the same ATT from the same microdata. The cross–national structure of ICILS, evaluated system by system, additionally allows the severe test to be applied repeatedly under comparable conditions, so that the paper can discuss whether robustness itself varies systematically across education systems and data–collection contexts, which is of direct relevance to researchers using comparable large–scale assessment or administrative infrastructures in other national settings, including Italy.

3. Methodology

MOMP operationalises each discretionary modelling decision as a "trigger": a categorical or discretised variable indexing one admissible option within a decision (for instance, nearest–neighbour versus radius matching, alternative caliper widths, alternative operationalisations of the socio–economic covariate, or alternative variance estimators). The set of triggers defines a combinatorial specification space; in the present application this space comprises several thousand distinct, individually defensible analytic pipelines.

Each pipeline in the enumerated space is run end to end, from matching to ATT estimation, and the resulting distribution of estimates is decomposed using variance–based sensitivity measures, namely Sobol first–order (S_1) and total–order (S^T) indices, which quantify how much of the variance in the estimated causal effect is attributable to each modelling choice individually and through its interactions with the others. This step formalises, with a quantitative variance decomposition, an exercise that applied researchers usually perform only informally through ad hoc robustness checks, which typically vary one choice at a time and cannot capture interaction effects between choices, a limitation that has a direct geometric interpretation, since one–choice–at–a–time exploration only samples a low–dimensional subset of a genuinely high–dimensional specification space.

The paper positions this exercise within Saltelli four–level hierarchy of sensitivity auditing (SAUD), moving from uncertainty analysis of the outcome, through sensitivity analysis of its drivers, to a discussion of the normative and framing choices – such as treatment definition and estimand selection – that condition the exercise but are deliberately kept outside the enumerated parameter space because they are theoretical rather than procedural decisions. The Average Treatment effect on the Treated is retained throughout as the target estimand, consistent with a behavioural rather than a programme–assignment framing of social media use.

4. Results and discussion

As the companion empirical analysis is still in progress, this abstract does not report numerical findings. The methodological contribution can nonetheless be characterised in terms of what the full MOMP analysis is designed to deliver. First, a variance decomposition will attribute the observed spread in ATT estimates across the enumerated specification space to specific analytic decisions, distinguishing choices that are largely inconsequential for the conclusion from those that are decisive. Second, applying the same procedure separately to each education system will make it possible to compare robustness across systems, rather than assuming that a single global robustness verdict applies uniformly to all of them. Third, and most directly relevant to the session's methodological focus, the analysis is designed to discriminate rather than simply to confirm or reject: MOMP is expected to certify the causal estimate as robust in some education systems, while flagging others as sensitive to specification choices that a conventional single–pipeline analysis would not have revealed. This discriminating behaviour is treated as evidence that the test is a genuinely severe one, in Popper and Mayo's sense, rather than a routine robustness check designed to reassure rather than to probe.

The discussion will also address the practical trade–off, already visible in the design phase of the companion study, between an optimisation logic that seeks a single defensible matching specification and a full–space exploration logic that treats the specification itself as an object of study; reconciling these two logics is part of the methodological contribution of the paper.

5. Conclusions

The paper argues that embedding a systematic, quantitative exploration of the modelling–choice space into counterfactual education policy evaluations offers a more severe and more transparent test of causal claims than conventional, informally chosen robustness checks. Modelling of the Modelling Process, grounded in variance–based Global Sensitivity Analysis and framed through the philosophy of severe testing, is proposed as a general instrument for this purpose, applicable beyond the specific ICILS/PSM illustration to any counterfactual evaluation built on a non–trivial chain of analytic decisions. For the education policy research community, the approach offers a way to report

not only a causal estimate but the extent to which that estimate depends on defensible choices that could reasonably have been made differently, information that is directly relevant to the credibility of evidence used to inform policy design and evaluation with large–scale assessment or administrative data infrastructures such as ICILS or INVALSI.

Keywords: Global Sensitivity Analysis; Severe Testing; Propensity Score Matching; Modelling of the Modelling Process

SESSION 2. WHAT MAKES A SCHOOL WORK? TEACHERS AND SCHOOL PRINCIPALS AS KEY ACTORS IN THE EDUCATIONAL SYSTEM

ORGANIZERS: UNIVERSITY OF PADOVA – UNIVERSITY OF MILAN "LA STATALE"
COORDINATORS: LORENZO ROCCO – MORIS TRIVENTI
4TH NOVEMBER: 4.30 P.M. – 6.30 P.M. {AULA A1.1 – RESEARCH 5}

Student and residential mobility from Southern Italy to the Centre-North: a comparison over the last ten years

Massimo Attanasio – Francesca Licari – Samuele Longo

Introduction

Internal mobility is a structural component of Italy's demographic history and reflects the deep economic, occupational and social asymmetries between the different areas of the country. After the Second World War, internal movements accompanied the processes of industrialization and urbanization, mainly taking the form of large-scale migration from Southern Italy to the industrial regions of the Centre-North. From the 1970s onwards, the intensity of these flows gradually declined, but the South-North route continued to represent one of the main axes of Italian mobility, within a framework marked by both continuity and change in migrants' profiles, motivations for moving and geographies of mobility (Bonifazi and Heins, 2000). Over the last ten years, just under 1 million 267 thousand individuals moved from Southern Italy to the Centre-North, compared with 663 thousand movements in the opposite direction, resulting in a negative balance for Southern Italy of more than 603 thousand residents (Istat, 2025).

Internal migration should be interpreted in close connection with territorial development gaps and with the unequal distribution of opportunities offered by local labour markets. In particular, movements from Southern Italy to the Centre-North retain a clear labour-related connotation: people mainly migrate towards areas where employment opportunities are greater, professional prospects are broader, and the productive system has a stronger capacity to absorb labour.

Particular attention should be paid to young people and, above all, to highly qualified young people. In recent years, mobility from Southern Italy to the Centre-North has increasingly taken the form of a selective transfer of human capital, with demographic and economic effects on both areas of origin and destination. The educational selectivity of migration trajectories shows how mobility contributes to reshaping the geography of skills and opportunities in the country (Basile et al., 2025). This is particularly evident among young graduates: in the period 2019-2023, Southern Italy recorded a net loss of around 94 thousand young graduates aged 25-34 towards the Centre-North, concentrated mainly in Campania and Sicily (Istat, 2025).

This dynamic increasingly appears to be an anticipated and hardly reversible phenomenon: migration does not occur only after entry into the labour market, but often tends to emerge already during university education, increasing the likelihood that the study path will translate into stable settlement in the areas of destination. According to the SVIMEZ–Save the Children report (2026), in the academic year 2024/2025 almost 70 thousand students from Southern Italy, out of approximately 521 thousand, were enrolled in universities in the Centre-North, accounting for more than 13% of the total; this share rises to 21% in STEM courses. Campania and Sicily alone generate almost half of the outgoing flow, while Lombardy, Emilia-Romagna and Lazio are confirmed as the main areas of attraction. The choice of university therefore appears to be closely linked to the opportunities offered by the territories of destination and may anticipate subsequent stabilization paths: three years after graduation, 88.5% of employed graduates who obtained their degree from a

university in the Centre-North work in the same macro-area, whereas among those who graduate in Southern Italy the share employed in their areas of origin remains below 70% (SVIMEZ–Save the Children, 2026).

The proposed analysis compares the trajectories of internal mobility with those of student mobility and highlights a substantial overlap between them. This comparison suggests that student movements should be interpreted not only as the outcome of the search for a better educational offer, but also — and above all — in light of the socio-economic and occupational characteristics of the destination areas. Student mobility would thus represent the first stage of a broader migration path, in which the decision to enrol in a university in the Centre-North anticipates possible stable settlement in areas characterized by more dynamic labour markets, greater professional opportunities and better prospects for labour-market integration.

The data used in the research are microdata from two sources: for the mobility of the resident population, the analysis refers to registrations and deregistrations due to changes of residence produced by Istat, concerning both Italian citizens and foreign nationals. For student mobility, instead, microdata from the National Student Register of the Ministry of University and Research were used.

The aim of the research is to compare flows of the resident population with those of student mobility for the years 2015 and 2021. The analysis focuses on movements out of the regions of Southern Italy towards the Centre-North, whose regions include universities that most strongly attract students from the South.

The geography of residential mobility from Southern Italy to the Centre-North

The analysis of residential mobility reveals substantial stability in the composition of flows between 2015 and 2021. In both years, a significant share of movements occurs within the same region of origin. In 2015, the highest levels of intra-regional mobility are observed in Sardinia, Campania, and Sicily, accounting respectively for 76.4%, 70.5%, and 69.3% of total movements. In 2021, the pattern remains largely unchanged: Sardinia 77.5%, Campania 70.4%, and Sicily 68.4%. By contrast, Basilicata records the lowest share of intra-regional movements, at 39.5% in 2015 and 38.1% in 2021, likely also reflecting the region's smaller demographic and territorial size (Table 1).

Among the Centre-North destinations considered, Lombardy confirms its role as one of the main areas of attraction. In 2015, the highest shares are recorded for Calabria, Basilicata, and Puglia, with 11.7%, 9.7%, and 9.2% of total movements, respectively. In 2021, Lombardy's role strengthens slightly for Calabria and Basilicata, reaching 12.6% and 10.7%, respectively, while Puglia remains stable at 9.2%. For Sicily as well, Lombardy remains the predominant destination among those considered, with its share increasing from 8.7% to 9.0%. Lazio, by contrast, assumes a specific role for Abruzzo and Molise, also due to geographical proximity. The share of movements directed to this region declines from 8.3% in 2015 to 7.1% in 2021, while still retaining the main destination among those considered.

Overall, between 2015 and 2021 no radical changes emerge in migration trajectories. Intra-regional mobility remains high, particularly in larger or insular contexts, while among extra-regional destinations the attractive role of Lombardy—and, to a lesser extent, Emilia-Romagna—is consolidated. Lazio slightly reduces its relative weight, while remaining central for Abruzzo and Molise.

Table 1a Internal migration, year 2015

	Region of destination (row %)								
Region of origin	Same region	Emilia-Romagna	Lazio	Lombardy	Piedmont and V. D'Aosta	Tuscany	Veneto	Other region	Total
Abruzzo and Molise	65.5	3.9	8.3	4.6	1.6	1.7	1.6	12.9	31763
Basilicata	39.5	7.8	7.6	9.7	4.4	4.4	2.0	24.6	7348
Calabria	53.1	5.6	7.5	11.7	4.6	3.3	2.2	12.0	33351
Campania	70.5	4.0	5.6	5.6	1.8	3.1	1.6	7.7	124916
Puglia	59.0	6.6	5.2	9.2	3.1	2.3	2.7	11.9	53819
Sardinia	76.4	2.4	4.0	5.4	2.8	2.1	1.4	5.5	29077
Sicily	69.3	3.6	3.4	8.7	3.4	2.4	2.4	6.8	85613
Total	242715	16117	19641	27048	9992	9772	7285	33317	365887

Table 1b Internal migration, year 2021

	Region of Destination (row %)								
Region of origin	Same region	Emilia-Romagna	Lazio	Lombardy	Piedmont and V. D'Aosta	Tuscany	Veneto	Other region	Total
Abruzzo and Molise	66.6	4.0	7.1	4.9	1.5	1.5	1.9	12.5	34930
Basilicata	38.1	8.4	7.2	10.7	3.6	4.2	2.9	24.8	8524
Calabria	52.7	6.3	6.3	12.6	4.2	3.0	3.2	11.7	36544
Campania	70.4	4.2	5.2	6.2	1.8	2.7	2.1	7.5	129992
Puglia	60.7	7.0	4.4	9.2	2.8	2.1	2.9	10.9	59939
Sardinia	77.5	2.4	3.1	5.5	2.4	1.7	1.7	5.7	31274
Sicily	68.4	4.3	3.0	9.0	3.2	2.3	3.0	6.7	91274
Total	260296	18815	18419	30728	10089	9289	9841	35000	392477

To assess the intensity of mobility in relation to the resident population, intra- and interregional mobility rates are considered (Table 2). Between 2015 and 2021, mobility levels increased: the overall intraregional mobility rate in Southern Italy rose from 0.60% to 0.66%, while the rate towards the six selected Centre-North regions increased from 0.44% to 0.49%. The highest intraregional mobility rates are observed in Basilicata and Calabria, reaching 0.98% and 0.93%, respectively, in 2021. Mobility towards the Centre-North is also more intense in Calabria and

Basilicata, with rates of 0.70% and 0.58%. Campania, Abruzzo and Molise, and Sicily show values close to 0.50%, while Puglia and Sardinia record lower levels.

Overall, intraregional mobility remains higher than mobility towards the six Centre-North regions, as expected, since it also includes moves related to geographical proximity, housing needs and family constraints. However, in Calabria, Campania, Puglia and Sicily, mobility towards the Centre-North has a relatively greater weight, indicating a stronger propensity to move towards the country's main attractive areas. The increase observed between 2015 and 2021 therefore confirms a moderate growth in mobility towards the main Centre-North poles, in line with the evidence from the origin-destination matrices.

Table 2. Mobility rates of the resident population, per 1000 inhabitants. Years 2015 and 2021

Region of Origin	2015		2021	
	Interregional Mobility Rate	Mobility Rate towards the 6 regions	Interregional Mobility Rate	Mobility Rate towards the 6 regions
Abruzzo and Molise	6.7	4.2	7.4	4.7
Basilicata	7.8	4.6	9.8	5.8
Calabria	8.0	6.0	9.3	7.0
Campania	6.4	4.7	6.9	5.1
Puglia	5.5	3.9	6.0	4.3
Sardinia	4.2	3.2	4.4	3.3
Sicily	5.3	4.1	6.0	4.7
Total	6.0	4.4	6.6	4.9

The comparison of standardized residuals[4] for 2015 and 2021 confirms the substantial stability of the main origin-destination trajectories and highlights some clearly recognizable territorial systems (Table 3). For Abruzzo and Molise, the strong relationship with Lazio is confirmed, with high and stable residuals in both years: 20.7 in 2015 and 20.5 in 2021. Calabria and Sicily, by contrast, show a pattern more clearly oriented along long-distance South-North trajectories. In these cases, positive residuals are mainly concentrated towards Lombardy and Piedmont-Valle d'Aosta, suggesting mobility patterns less driven by geographical proximity and more directed towards Northern areas with stronger economic and employment attractiveness. Lombardy continues to play a particularly

[4] Standardized residuals are obtained from a main-effects log-linear model, estimated separately for 2015 and 2021. In other words, residuals make it possible to identify the origin-destination combinations for which observed flows are higher or lower than expected under the assumption of independence. A positive residual indicates that the observed flow between a specific region of origin and a specific region of destination is higher than expected, while a negative residual indicates a lower-than-expected flow.

important role: the residual increases for Calabria, from 7.9 to 9.0, while for Sicily it declines from 21.7 to 17.6, remaining nevertheless very high.

Table 3. Comparison of standardized residuals for the years 2015 and 2021

residuals (2015)	Emilia-Romagna	Lazio	Lombardy	Piedmont and V. D'Aosta	Tuscany	Veneto	Other region
Abruzzo and Molise	-5.2	20.7	-19.5	-12.2	-10.5	-6.1	20.4
Basilicata	-0.1	-5.4	-8.9	-1.8	-1.1	-6.6	16.7
Calabria	-3.8	0.2	7.9	7.5	-4.0	-6.5	-3.6
Campania	3.2	15.3	-11.6	-14.7	16.9	-2.7	-3.3
Puglia	12.4	-12.1	1.5	-2.9	-12.1	3.3	5.9
Sardinia	-6.5	2.2	1.8	10.8	3.2	-0.2	-6.4
Sicily	-5.5	-19.6	21.7	17.3	-0.4	12.0	-15.6
residuals (2021)	Emilia-Romagna	Lazio	Lombardy	Piedmont and V. D'Aosta	Tuscany	Veneto	Other region
Abruzzo and Molise	-6.3	20.5	-19.2	-11.4	-10.7	-6.9	22.9
Basilicata	-1.0	-4.6	-9.0	-4.6	-0.6	-7.1	18.9
Calabria	-2.5	-2.1	9.0	6.3	-3.8	-3.0	-4.9
Campania	-1.2	17.9	-9.0	-11.9	14.6	-2.0	-3.7
Puglia	15.7	-10.8	1.0	-3.1	-9.6	-0.6	2.4
Sardinia	-7.6	-0.2	1.9	8.7	2.3	-0.2	-1.8
Sicily	6.97	-8.23	6.67	14.28	7.85	6.59	-18.82

The geography of student mobility from Southern Italy to the Centre-North

Table 4a Student Mobility from Mezzogiorno to Center-North Italy, ay 2015–2016

Student Origin–Destination, ay 2015–2016 (Row %)									
	Region of Destination								
Region of Origin	Same Region	Emilia-Romagna	Lazio	Lombardy	Piedmont & Aosta Valley	Tuscany	Veneto	Other Regions	Total
Abruzzo & Molise	62.5	7.3	8.1	4.5	2.1	2.2	1.2	12.00	9976
Basilicata	28.7	8.1	10.5	6.4	4.7	8.0	0.6	33.0	3175
Calabria	61.0	4.2	8.7	5.9	2.6	4.6	0.5	12.5	10973
Campania	85.8	1.2	5.7	2.0	0.9	1.3	0.3	2.9	36520
Puglia	64.1	6.2	5.4	5.6	4.3	2.4	1.0	11.1	22541
Sardinia	80.8	3.1	3.1	3.0	3.4	3.0	1.1	2.5	7622
Sicily	70.1	4.5	5.3	5.0	4.5	4.2	1.2	5.3	25683
Total	83777	4681	6982	4790	3331	3284	915	8730	116490

Table 4b Student Mobility from Mezzogiorno to Center-North Italy, ay 2021–2022

Student Origin–Destination, ay 2021–2022 (Row %)									
	Region of Destination								
Region of Origin	Same Region	Emilia-Romagna	Lazio	Lombardy	Piedmont & Aosta Valley	Tuscany	Veneto	Other Regions	Total
Abruzzo & Molise	57.9	9.0	11.3	5.2	1.9	2.4	1.6	10.7	10733
Basilicata	25.5	10.4	10.6	5.9	4.2	6.2	0.7	36.4	3348
Calabria	63.2	5.0	8.3	5.4	2.1	3.3	0.9	11.9	11996
Campania	84.9	1.6	6.1	2.4	0.6	0.9	0.3	3.2	40189
Puglia	68.3	6.5	5.5	5.8	2.9	1.9	1.2	7.9	26071
Sardinia	80.0	3.9	3.1	3.5	3.0	2.1	1.8	2.6	8559
Sicily	75.8	5.1	4.5	4.7	3.2	2.5	1.3	3.2	30170
Total	96252	6111	8055	5604	2820	2666	1265	8293	131066

Tables 4a and 4b report data on student mobility for 2015 and 2021. The flows were calculated by including first-year undergraduate students; bachelor's degree graduates in 2014–15 (belonging to the cohorts of

students who enrolled in 20–2013, 2013–14, and 2014–15), who enrolled in a master's degree programme in 2015 at universities in the same region, in universities located in the six regions considered in this study, or in other regions. In 2015, the region receiving the largest number of students was Lazio, followed by Lombardy and Emilia-Romagna, while Emilia-Romagna overtook Lombardy in 2021. In both years, Lazio was almost always the region with the greatest capacity to attract students from regions whose distance from Rome is not particularly great. Indeed, students from the two major islands were distributed more or less evenly across all the regions, with the exception of Veneto, while Puglia retained Emilia-Romagna as its destination, following the Adriatic corridor.

Tab. 5 "Out-of-Region" Mobility Rates and Mobility Rates "to the Six Regions" in A.y 2015–16 and 2021–22 (%)

Mobility Rates	Ay 2015–16		Ay 2021–22	
Region of Origin	Out-of-Region	To the Six Regions	Out-of-Region	To the Six Regions
Abruzzo & Molise	37.5	25.6	42.1	31.4
Basilicata	71.3	38.3	74.5	38.1
Calabria	39	26.6	36.8	25
Campania	14.2	11.3	15.1	11.9
Puglia	35.9	24.8	31.7	23.9
Sardinia	19.2	16.7	20	17.5
Sicily	29.9	24.6	24.2	21.1
Total	28.1	20.6	26.6	20.2

From Table 5, it emerges that overall in Southern Italy the attractiveness of the six regions considered is stable for all regions, with the exception of Abruzzo and Molise, which show an increase of 6 percentage points from 2015 to 2021, while in Sicily a decrease is observed in the same years. Furthermore, a considerable difference is noted between "Out-of-Region" and "Towards the Six Regions" in regions where there are no large universities, such as Abruzzo, Calabria, and Basilicata. This is since most students go to universities located in neighboring regions. By now decomposing the share of mobility at enrolment with that at the master's level, an increase in percentage is observed: from 13.9% in 2015 to 16.5% in 2021, with large variability, in fact values always between 2 and 3% for Basilicata, where mobility occurs mostly at the bachelor level, and high values in Campania, with values equal to 25% in 2021.

Tab. 6 Comparison of the Standardized Residuals between the 2015–16 and 2021–22

residuals (2015-16)	Emilia-Romagna	Lazio	Lombardy	Piedmont & Aosta Valley	Tuscany	Veneto	Other Regions
Abruzzo & Molise	8.2	0.2	-4.3	-7.7	-7.9	1.4	5.9
Basilicata	-3.9	-7.2	-6.5	-5.1	1.7	-4.2	17.0
Calabria	-6.0	1.3	0.8	-7.6	3.7	-5.4	6.9
Campania	-10.5	28.9	-1.6	-9.3	-2.6	-3.8	-8.4
Puglia	6.8	-12.0	1.8	5.0	-9.4	-0.2	7.2
Sardinia	1.6	-4.4	1.1	9.3	6.7	7.2	-10.1
Sicily	1.9	-6.7	4.8	12.0	10.7	5.1	-14.8

residuals (2021-22)	Emilia-Romagna	Lazio	Lombardy	Piedmont & Aosta Valley	Tuscany	Veneto	Other Regions
Abruzzo & Molise	6.0	5.0	-6.1	-8.4	-4.7	0.4	2.3
Basilicata	-4.8	-9.1	-10.2	-3.6	1.4	-6.1	24.6
Calabria	-6.2	-0.8	-2.3	-5.7	2.7	-4.4	11.4
Campania	-12.9	27.5	-0.3	-11.0	-4.5	-6.3	-4.4
Puglia	6.3	-11.0	5.3	3.3	-5.0	0.2	2.1
Sardinia	2.0	-6.7	1.2	11.2	4.8	10.8	-9.4
Sicily	7.0	-8.2	6.7	14.3	7.9	6.6	-18.8

The most attractive regions for Sicily and Sardinia are Piedmont, Tuscany, and Veneto, although from 2015 to 2021 the intensity of attraction towards Tuscany decreased for both regions, while that towards the other two regions increased for both. For Campania, the most attractive region is Lazio, and the intensity of attraction slightly decreases between the two years. Lombardy became more attractive for Puglia and Sicily in 2021 compared to 2015. Emilia-Romagna is the most attractive region for students coming from Puglia and Abruzzo and Molise in 2015; however, in 2021 the origin region with the largest standardized residual towards Emilia-Romagna becomes Sicily, increasing from 1.89 in 2015 to 6.97 in 2021.From Basilicata, none of the six destination regions analysed appears to be particularly attractive; the highest residual is associated with “Other regions”, because historically students tend to go to neighbouring regions, Campania and Puglia, to the universities of Bari, Salerno, or Naples, depending on distance. Other few flows show opposite trends.

Conclusions
The analysis of population and student mobility flows reveals recognizable patterns that remain broadly consistent across the years considered. Furthermore, the residual analysis highlights a strong agreement between population and student flows, both in terms of direction (sign) and, in several cases, in terms of their quantitative magnitude, in both 2015 and 2021. Further investigation incorporating the additional variables available in the dataset could enrich the analysis and provide a deeper understanding of the factors underlying the observed mobility dynamics.

Keywords: Student mobility; population mobility; Italian North-South divide; Pearson residuals

School Leadership and the Conditions for Effective Schools: Global Evidence from the GEM Report PEER Profiles

Anna Cristina D'Addio – Daniel April

What makes a school work depends not only on resources, curricula or assessment systems, but also on the people who translate policy into everyday educational practice. This paper examines school principals and teachers as key actors in effective schooling through a global analysis of the 2024 GEM Report Profiles for Enhancing Education Reviews (PEER) on school leadership. Produced by UNESCO's Global Education Monitoring Report, the PEER profiles document education laws and policies in relation to SDG 4 and the thematic focus of the GEM Report. They are important because they provide a comparable, evidence–based overview of how education systems are organized across countries, making it possible to identify global patterns, highlight policy gaps, support mutual learning and inform national and international policy dialogue.
Drawing on evidence from 211 education systems, the paper explores how countries define, regulate and support school leadership in primary and secondary education, with particular attention to principals' professional roles, teacher collaboration, teacher evaluation, leadership preparation and school governance. The analysis shows that school leadership is increasingly professionalized, yet unevenly developed across systems. While many countries have introduced professional standards for school principals, these frameworks often emphasize vision–setting, school management and instructional improvement more than collaborative or distributed leadership. Principals are frequently entrusted with teacher evaluation and pedagogical supervision, positioning them as central mediators between system–level expectations and classroom practice. However, selection and preparation pathways remain highly variable: teaching experience is often the main requirement for appointment, while management experience and structured induction are less consistently required. Although most systems provide some form of in–service training, fewer offer systematic pre–service preparation, and only a minority of leadership training programmes cover the full range of instructional, transformational, staff development and collaborative leadership dimensions.
The paper argues that effective schools require a broader conception of leadership: one that combines trust in principals, adequate professional preparation, meaningful teacher participation and inclusive governance involving parents, students and communities. The findings suggest that school improvement depends less on individual heroic leadership than on enabling conditions for shared responsibility, professional dialogue and locally grounded decision–making. By comparing global policy patterns, the paper contributes to debates on what makes schools work and identifies policy levers for strengthening principals and teachers as co–constructors of educational quality.

Keywords: School leadership; Professional standards; Teacher evaluation; Teacher collaboration; School governance; Leadership preparation; Education policy; PEER profiles; GEM Report; Comparative education

Women Leading for Learning: Gender Gaps in School Leadership and Their Consequences for Educational Quality

Anna Cristina D'Addio

Despite comprising the vast majority of the global teaching workforce, women remain profoundly underrepresented in school leadership positions worldwide. This paper draws on the 2025 Gender Report of UNESCO's Global Education Monitoring (GEM) Report — "Women Lead for Learning" — to examine the structural dimensions of this leadership gap and its consequences for educational quality, teacher professional development, and student outcomes.

Drawing on comparative data from international large–scale assessments — including TALIS, PASEC, SEA–PLM, and PISA — as well as national datasets from over 70 countries, the paper documents a persistent 20–percentage–point gender gap in school leadership at the secondary level globally. In sub–Saharan Africa, the disparities are especially acute: as few as 5–16% of primary school principals are women across francophone African countries, with consequences not only for equity but for school effectiveness.

The paper addresses three interrelated questions.

First, what do we know empirically about gender differences in school leadership practices? Evidence from TALIS data and studies across multiple regions suggests that female principals devote comparatively more time to curriculum matters, student and parent interactions, and collaborative school cultures — practices linked to instructional quality.

Second, what are the measurable effects of female school leadership on student learning? Analysis of PASEC data reveals that schools under female leadership in Benin, Madagascar, Senegal, and Togo achieved learning outcomes equivalent to an additional year of schooling compared to male–led schools. Similarly, SEA–PLM data show children in female–led primary schools in Cambodia, Lao PDR, and Myanmar gained four to six additional months of learning.

Third, what structural and societal barriers prevent talented women from reaching school leadership positions, and what policies have proven effective in addressing them?

The paper argues that the underrepresentation of women in school leadership is not a supply–side problem rooted in women's lack of ambition or capability, but a demand–side failure driven by institutional biases, opaque recruitment processes, inadequate mentoring, and entrenched societal norms. It presents a typology of barriers — societal, institutional, and developmental — and reviews evidence on policy interventions, including affirmative action provisions, structured mentoring programmes, leadership training initiatives, and awareness campaigns. The paper concludes with a call for systems–level monitoring of the leadership pipeline, transparent and bias–aware recruitment practices, and targeted investment in women's leadership development as strategic levers for improving school quality.

The implications for the discussion on leadership are direct: gender–equitable school leadership is not merely an equity imperative but an educational effectiveness imperative. Understanding who leads schools — and under what conditions effective leadership is fostered irrespective of gender — is central to the question of what makes a school work.

Keywords: School leadership; Gender equity; Instructional leadership; Principal effectiveness; Comparative education; SDG 4

How Are School Track Recommendations Generated? Survey Evidence from Italian Teachers

Giorgio Brunello – Clementina Crocè – Pamela Giustinelli – Lorenzo Rocco

Introduction

Teacher recommendations can shape students' educational trajectories in tracked education systems. This paper studies how Italian lower–secondary teachers translate subjective beliefs about student outcomes into track recommendations. Italy provides a relevant setting because upper–secondary education is characterised by curricular tracking, a feature it shares with most OECD countries (Betts, 2011), and lower–secondary teachers are required by law to provide a formal recommendation for one or more upper–secondary tracks to final–year students and their families. The recommendation is non–binding, and families retain the final enrolment decision. It is nevertheless salient, with 56 percent of students enrolling in the recommended track (Fondazione Agnelli, 2021). Track choice is consequential: post–secondary pathways differ sharply across upper–secondary diplomas, and even among university graduates labour–market returns are higher for students who attended academic tracks than for those who attended technical or professional tracks (Agarwal et al., 2021).

Existing work documents systematic differences in track recommendations across student groups (Van Huizen et al., 2025) and, in some settings, links such differences to teacher stereotypes (Carlana, 2019; Carlana et al., 2022). Such gaps are policy–relevant, but they do not by themselves reveal how recommendations are produced. A gap may arise because teachers forecast different track–contingent outcomes for otherwise similar students, because they place different weights on the same forecasts, or because student characteristics enter advice through considerations not captured by the elicited outcomes. These mechanisms point to different interventions, so understanding the recommendation process is a prerequisite for deciding how to respond to observed gaps.

This paper estimates how teachers map those beliefs into recommendations. Our first question is whether teachers take long–run outcomes into account at all – a policy–relevant issue because the recommendation can influence choices with lasting consequences – or instead focus mainly on short–run school success, namely preparedness and regular completion. We then examine whether uncertainty changes the weight attached to long–run expectations.

The analysis answers the positive question of whether and how long–run forecasts enter recommendations; determining the socially optimal weight would require an explicit welfare benchmark and evidence on realised outcomes.

The paper contributes to four strands of literature. First, it relates to work on teacher characteristics and expectations (Hanushek and Rivkin, 2006; Bacher–Hicks and Koedel, 2023; Gershenson et al., 2016; Hill and Jones, 2021; Papageorge et al., 2020; Sabarwal et al., 2022; Brunello et al., 2025). Second, it contributes to research on teacher track recommendations and educational inequality (Carlana, 2019; Carlana et al., 2022; Bernardi et al., 2014; Bach, 2023; De Ree et al., 2023; Van Huizen et al., 2025; Falk et al., 2025; Van Maarseveen, 2024; Biroli et al., 2025). Third, it draws on literature on subjective–expectations and choice under uncertainty (Manski, 2004; Blass et al., 2010; Wiswall and Zafar, 2015; Giustinelli, 2016; Conti and Giustinelli, 2025; Meango and Girsberger, 2024). Fourth, it connects to work on preferences, perceived ability, and returns in educational choice (Wiswall and Zafar, 2015; Boneva and Rauh, 2020; Boneva et al., 2022; Patnaik et al., 2022; Brunello et al., 2026).

The empirical analysis uses a vignette experiment administered to approximately 1,000 teachers in a stratified national sample of Italian lower–secondary schools. Each teacher evaluated two hypothetical students, Paolo and Francesca, whose migration background, parental occupations, self–confidence, academic profile, and university aspirations were independently randomised. For each student and track, teachers reported probabilistic beliefs about outcomes and a probability of recommending the track. Random assignment makes the student attributes orthogonal to teacher

characteristics and generates within–teacher variation in both beliefs and recommendations; the structural interpretation of the estimated belief weights is supplied by the analytic framework and its identifying assumptions.

Two robust results emerge from the main analysis. First, teachers are forward–looking: the long–run job belief enters recommendations positively and significantly. Yet the short–run beliefs receive larger weights, especially the belief that the student will be adequately prepared. Second, uncertainty attenuates the long–run belief weight.

Data

The evidence comes from a survey fielded between January and June 2025. Schools were drawn at random from strata defined by region and municipality type. Invitations were sent by email and followed by a telephone call. Schools that declined were replaced by a new random draw from the same stratum until fieldwork targets were reached. Sampling, recruitment, and fieldwork were conducted by a professional survey firm. Of 390 contacted schools, 151 participated, a response rate of approximately 39 percent. Observable characteristics of participating and non–participating schools are very similar: two–sample tests for teacher counts and composition, student counts, gender composition, and migration background yield no statistically significant differences.

In total, 1,002 teachers submitted complete questionnaires. In the full respondent sample, 78.2 percent are women, the average age is 47.5 years, and average teaching experience is 15.7 years. Approximately 95 percent teach eighth–grade classes and 82 percent hold an open–ended contract. Around 12 percent serve as the school orientation coordinator and 5 percent as the class–level orientation lead.

The survey contained a vignette experiment. Each teacher evaluated two hypothetical but realistic eighth–grade students choosing among five upper–secondary tracks: academic–classical studies, academic–scientific studies, other academic studies, technical education, and professional education, including regional vocational education and training in the survey label. One vignette student was named Paolo and the other Francesca, and their order was randomised. Within each vignette, the items of every student dimension were assigned independently of the others. The randomised dimensions were migration background, academic profile, self–confidence, parental occupations, and university aspirations. Gender was signalled by the student name. The independent randomisation generated 192 possible profiles for each name 192^2 possible paired assignments across Paolo and Francesca.

For each vignette and every potential track, teachers used 0–100 sliders to report outcome probabilities. The main analysis uses three belief measures: (1) whether the student would be adequately prepared for the track; (2) whether the student would complete upper–secondary school on time without repeating a year, dropping out, or changing track; and (3) whether the student would have a stable and satisfactory job at age 30. The good–job belief integrates the teacher's expectation that the student attends university with conditional expectations about a stable and satisfactory job at age 30.

Teachers then assigned a probability of recommending each track. Unlike the outcome beliefs, these five recommendation probabilities were constrained to sum to 100, so each answer is a compositional distribution over tracks.

The analysis sample contains the 732 teachers with complete recommendation and belief responses for both vignettes and all five tracks.

Methodology

The paper uses a subjective expected–welfare representation of the recommendation problem. For student i and track s, teacher j evaluates the student's expected welfare from that track. The teacher's subjective probability that an outcome occurs enters expected welfare through the value the teacher assigns to that outcome for student i. The observed object is not a realised track choice but a stated recommendation probability. Teacher j's latent utility from recommending track s is the expected welfare from that track plus information that is visible to the teacher but omitted from the vignette

and a residual uncertainty term. The first unobserved component captures track–specific considerations that the teacher can imagine but the analyst does not observe. The second captures residual uncertainty. With Type–I extreme–value errors, stated recommendation probabilities take a multinomial–logit form, and the estimating equation uses log–odds relative to a reference track.

The empirical specification sets other academic studies as the reference category. The dependent variable is the log–odds of recommending a track relative to the reference track. The key explanatory variables are belief differences relative to that reference track for preparedness, regular completion, and a good job at age 30. The coefficients for these variables are belief weights: they measure how teachers translate a track's belief advantage over the reference track into a recommendation–odds advantage. Teacher–by–track fixed effects absorb stable teacher–specific preferences for each track. Both dependent variable and belief differences are standardised, and standard errors are clustered by teacher. Richer specifications add student–profile controls and interactions that allow belief weights to vary with uncertainty.

Random assignment of student traits makes the vignette attributes orthogonal to teacher characteristics and creates within–teacher variation in beliefs and recommendations. Each teacher evaluates two independently drawn profiles, and teacher–by–track fixed effects absorb any stable tendency to favour or disfavour a track. The beliefs themselves are post–randomisation outcomes, however, so randomisation alone does not make the recommendation–on–belief regression causal. Interpreting the coefficients as structural belief weights additionally requires that, conditional on randomised profile controls and teacher–by–track fixed effects, omitted student–track considerations are not systematically correlated with relative beliefs and that measurement error in the elicited probabilities is adequately handled.

The paper also explores belief precision. Numerical expectation surveys often produce focal 50–percent answers when respondents are unsure rather than expressing a precise probability (Bruine de Bruin et al., 2000; Manski, 2004). The paper uses focal 50–percent responses in the long–run items to construct an uncertainty indicator. The augmented specification allows the weight on the good–job belief to differ when this uncertainty indicator is active. The prediction is that imprecise long–run forecasts should receive less weight.

Results and discussion

The main analysis yields two robust results. First, teachers are forward–looking: beliefs about a stable and satisfactory job at age 30 enter the recommendation decision positively and significantly. Second, short–run school–success beliefs receive larger weights, especially the belief that the student will be adequately prepared. The estimates therefore show that teachers do use expected labour–market outcomes when recommending tracks, but they place larger weights on preparedness and regular completion.

In the baseline specification, all three belief weights are positive and precisely estimated. In the specification with teacher–by–track fixed effects, the preparedness weight is 0.343, the regular–completion weight is 0.208, and the good–job weight is 0.132. Student controls reduce the three estimates to 0.205, 0.145, and 0.108, but preserve their ordering.

The finding establishes that long–run labour–market prospects are part of the recommendation rule; it does not establish that their estimated weight is socially too low or too high.

The long–run component is fragile when teachers are unsure. For observations without the uncertainty indicator, the long–run weight is 0.176. The interaction between uncertainty and the good–job belief is –0.161, leaving an implied weight of 0.015 under uncertainty; this sum is not statistically different from zero. The short–run interactions are statistically insignificant. The pattern is consistent with teachers discounting distant forecasts when they regard them as imprecise.

We discuss two interpretations of the greater weight assigned to short–run beliefs. One interpretation is informational. Whether a student is prepared for a track and likely to complete it regularly is closely connected to teachers' daily experience. Labour–market outcomes at age 30 are distant, depend on further education and macroeconomic conditions, and are harder to forecast. The

uncertainty results support this interpretation: teachers who signal imprecision through focal responses largely stop using the long–run belief.
A second interpretation concerns the guidance environment. Only 27.9 percent of teachers report meetings with labour–market experts as part of the counselling process, compared with much more frequent contact with upper–secondary teachers and students. Better information about local employment prospects could improve both the quality of long–run beliefs and the confidence with which teachers use them. Such information would need to complement, rather than replace, assessments of academic fit and student preferences.
Taken together, the results show that teachers are not completely myopic. Expected labour–market outcomes at age 30 enter their recommendations. Nevertheless, recommendation probabilities remain anchored primarily in short–run school success, and the long–run component is especially fragile when teachers are uncertain.

Conclusions

The paper provides survey–experimental evidence on how Italian lower–secondary teachers produce upper–secondary track recommendations. Teachers evaluate two independently randomised student profiles, state track–specific beliefs about short– and long–run outcomes, and allocate recommendation probabilities across five tracks.
The main analysis yields two robust conclusions. First, teachers are forward–looking: beliefs about a stable and satisfactory job at age 30 enter the recommendation decision. Second, short–run school–success beliefs receive larger weights, and uncertainty can reduce the implied long–run weight to nearly zero.
The analysis has limitations. Vignette responses may differ from recommendations issued for students whom teachers know personally, and the elicitation necessarily simplifies the information available in a real class council. The sample covers one wave of teachers in 2025, and the current data do not link vignette recommendations to subsequent enrolment or realised outcomes. The randomised profiles nevertheless provide a controlled environment in which beliefs and recommendations are measured on common probability scales.
These findings establish that teachers do take long–run outcomes into account, while placing greater emphasis on immediate school fit; they do not establish the optimal balance between horizons. Policies aimed at improving recommendations should therefore be tied to the diagnosed mechanism. Better labour–market information may improve the precision and salience of long–run beliefs, whereas recommendation gaps driven by different weights or residual considerations would call for different responses. Any such policy should preserve attention to preparation and student interests.

Keywords: Teacher recommendations; Subjective expectations; School tracking; Italy

The Hidden Teacher Effect: Instructional Continuity and Assessment Accuracy A Research Hypothesis on Fifth–Year Classes in Upper Secondary School

Claudia Califano

Schools today have increasingly refined tools for measuring learning — standardized tests, authentic–assessment rubrics, formative assessment systems — yet they systematically neglect the most pervasive variable in the entire evaluative process: the teacher, understood not as a neutral instrument of measurement but as a cognitive, relational and institutional agent who, to a non–negligible extent, helps construct that very measure. This role unfolds along dimensions that are hard to quantify: teachers' personal beliefs about students, the expectations they hold, the attitudes they bring to assessment, and the social images they build of different groups in the class. Rarely present in administrative data, these factors have been called the teacher's "hidden effects" — psychological,

cognitive and relational dynamics that shape everyday educational interactions and that can amplify or attenuate inequalities.

"Hidden" here does not mean statistically marginal but structurally invisible to the ordinary tools used to assess teaching quality and system fairness: a factor operating upstream of measurement, which measurement itself cannot capture except through empirical evidence documenting its pervasiveness.

Over the past thirty years, Da Re's studies and guidelines have helped define what constitutes coherent, quality assessment by teachers. While not addressing implicit bias directly, his work offers concepts for understanding how assessment is crossed by subjective dimensions and possible distortions, and how professionalism requires constant critical reflection on one's own frameworks of judgment — bridging evaluative pedagogy with recent evidence from the economics of education and social psychology.

Educational psychology shows that teacher expectations profoundly affect students' motivation, self–efficacy and engagement, producing genuine self–fulfilling prophecies (Rubie–Davies, 2024; Urhahne and Wijnia, 2021); the sociology of education has linked these processes to inequalities of social origin and to the stratification of school pathways. Rubie–Davies (2024) argues that teachers' *beliefs* about learning, assessment, curriculum and the purposes of education should be studied on a par with professional competencies, since they are a direct antecedent of daily instructional choices and of how students come to see themselves.

Unlike much of the literature on *expectations* and *bias* as mechanisms shaping instruction before any formal assessment, Wedenoja, Papay and Kraft (2022) isolate a different, complementary mechanism: relational continuity as a variable acting independently on the quality of the implicit assessment a teacher builds over time.

An observation from direct experience teaching Italian and Latin in upper secondary school points in the same direction: having the same teacher for several years in the same subject seems to affect students' confidence and learning progression noticeably. Students appear to experience a stable teacher presence as an anchor within a trusted relationship, contributing to more stable outcomes over time.

To what extent are students' educational trajectories a product of learning, and to what extent a product of attribution? This paper explores this question through a specific, under–studied variable: the continuity of the teacher–student relationship over time, as a factor acting independently on the quality of the implicit assessment a teacher constructs.

Data Used

The analysis draws on four strands of sources: school administrative data (internal grades, standardized test results, guidance recommendations, track enrollment), international literature, longitudinal studies, and control data. Administrative data made it possible to trace students' trajectories and link internal assessments to objective performance, though fragmentation across sources required considerable integration work even within the current year's fifth–year classes.

Wherever possible, information on teacher continuity within the same class and subject was drawn from the "15 May" documents of eight classes at the institution where I teach, and correlated with both the admission–board outcome and the final state–examination result — aimed at empirically assessing how continuity affects student outcomes in fifth–year upper–secondary classes.

The second source is the international literature on teacher–student continuity: Kraft, Papay and colleagues (Brown University / Tennessee Education Research Alliance, 2019–2022), using nearly a decade of Tennessee administrative data (grades 3–11); and Cistone and Shneyderman (2004) on looping in Florida (grades 2–5), with small but positive effect sizes (d = 0.16 reading, d = 0.21 mathematics).

Further material comes from recent studies such as Argentin (2026), from Ministry of Education data on staffing and retirements (a proxy for structural instability in the Italian system), from INVALSI's working papers on school and class variability (invalsiopen.it), and from the fourth national IARD

survey of Italian teachers (University of Milano–Bicocca, 2025), which supplies up–to–date data on stability and turnover. An earlier teacher survey using internationally validated questionnaires (Geven et al., 2021; Wang, Rubie–Davies and Meissel, 2018) measured performance expectations, perceptions of student ability, beliefs about the malleability of intelligence, and attitudes toward specific groups (e.g., students with a migration background or low socioeconomic status), allowing systematic measurement of teacher beliefs and expectations.

Together, these sources make it possible to relate teachers' beliefs, expectations and assessment practices to students' outcomes and choices, while accounting for socioeconomic context.

The literature on teacher expectations spans over half a century. Hattie's (2023) synthesis of nine meta–analyses yields an average effect size of d = 0.58 — substantial by educational–research conventions, though it averages over structural differences between subgroups. A more nuanced view comes from Jussim and Harber (2005): across thirty–five years of evidence, self–fulfilling prophecies exist but are generally modest, do not accumulate strongly over time, and tend to dissipate; effects for stigmatized groups fall between d = 0.2 and 0.3, with higher peaks under marked social vulnerability. The two positions are not contradictory but methodologically complementary: averaging conceals distribution, and it is in the distribution — the effect on specific subgroups — that the pedagogically relevant issue lies.

A second, often underestimated finding concerns inter–teacher variance: about 27% of the variance in teacher expectations is attributable to the class — that is, the teacher — rather than the student (Hollenstein, Brühwiler and König, 2024), and "high–" or "low–expectation" teachers keep this profile stable across years and cohorts, with an average outcome difference of d = 0.85 (Rubie–Davies and Hattie, 2024). Expectations, then, are not purely a rational response to observable student traits but also reflect stable cognitive structures of the teacher.

When expectations translate into formal decisions — grades, admission judgments, guidance — their effects amplify: the literature documents systematic bias by gender, origin and socioeconomic status in grading identical work (Sprietsma, 2013), and shows that teacher expectations predict subsequent track choice more strongly than standardized achievement does (De Boer, Bosker and Van der Werf, 2010). It is at this level — where individual judgment becomes an institutional decision with long–term consequences — that the question of this paper is situated.

Among longitudinal and control studies, a few merit particular mention. Gentrup, Lorenz, Kristen and Kogan (2020, *Learning and Instruction*) followed 1,026 first–year students in 64 German classes, finding that early teacher expectations predict year–end learning even after controlling for entry achievement — evidence of inaccurate, not merely self–confirming, expectation effects. Wang, Rubie–Davies and Meissel's (2018) review of 144 studies over thirty years found systematic bias by gender, ethnicity and socioeconomic status — expectations higher for girls in reading and boys in mathematics, and lower for ethnic–minority and low–income students — together with strong longitudinal stability in teachers' high– or low–expectation profiles. Rubie–Davies' (2024) narrative review in the *New Zealand Journal of Educational Studies* confirms the d = 0.85 gap between high– and low–expectation teachers, replicated in Poland (Szumski and Karwowski, 2019) and Germany (Knigge et al., 2016).

Methodology

The methodology proceeds in three steps. First, teachers' grading severity was estimated by comparing internal grades with standardized test scores for the fifth–year classes at the institution where I teach, isolating the subjective component of judgment and distinguishing stricter from more lenient graders relative to an external standard — a step limited by the impossibility of controlling for all contextual variables, notably class composition and past teacher changes. The national and international literature reviewed above provided the framework for situating this paper's focus: the impact of instructional continuity on learning outcomes, within the "hidden teacher effect."

To identify this effect independently of observable characteristics, (quasi–)exogenous variability in student–teacher assignment was considered, following a strategy used in recent studies (Brunello et

al., 2025; Sorensen and Ladd, 2024) — analyzing cases where assignment to classes and teachers is non–selective, reducing selection bias. Subjective expectations are further examined through probabilistic elicitation techniques, measuring the anticipatory component of teacher judgment and its impact on students' pathway choices.

Results and Discussion

Teachers' grading severity concretely affects students' trajectories: stricter graders tend to penalize students from disadvantaged backgrounds more heavily, with negative consequences for guidance recommendations and access to more selective pathways — suggesting that variability in grading severity can reinforce inequality, especially at transition points.

The data also suggest that instructional continuity has a positive effect not only on learning but on classroom climate and motivation. Personal observation confirms that a stable teacher presence fosters mutual familiarity, enabling instruction better tailored to students' needs — especially for those in situations of fragility or disadvantage. This aligns with Loeb, Kalogrides and Béteille (2012), who found in the United States that keeping the same teacher with the same class, especially in primary school, significantly improves learning and reduces inequality: students taught by the same teacher over several years outperform peers who change teachers frequently, benefiting from a more stable climate, greater mutual familiarity, and more targeted interventions — particularly for disadvantaged students — while teacher mobility can hinder achievement in more vulnerable schools.

Teacher expectations remain a key factor: high expectations are linked to better standardized–test results, motivation, self–efficacy and lower dropout risk; low expectations can generate self–fulfilling prophecies that limit opportunity and widen social inequality. Expectations also interact with socioeconomic background — disadvantaged students benefit most from high expectations and are most vulnerable to low ones. Implicit stereotyping plays a role too: at equal achievement, students with a migration background or low socioeconomic status receive less favorable recommendations, consistent with findings from other European countries (Burgess and Greaves, 2013; van Leest et al., 2021).

Overall, these unobservable dimensions appear to matter as much as variables traditionally seen as crucial — experience, qualifications, instructional continuity — suggesting they are a structural feature of the education system that should be integrated into models of teaching–effectiveness assessment and into education policy.

Assessment Practices as Determinants of Educational Trajectories

Assessment as a Performative Act

The evidence so far shows how teacher expectations shape classroom behavior and, through it, student outcomes. A further question is how expectations translate into formal assessment decisions with long–term effects. Here, school assessment is read not as measuring a pre–existing competence but as a performative act — in Austin's sense — a practice that does not merely register reality but helps produce it. Grades, judgments and guidance decisions shape how students see themselves, orient future choices, and determine access to differentiated resources: a grade is not a photograph of competence but a device that helps construct a student's school career.

Bias in Formal Assessment and School Guidance

Experimental studies using anonymous–versus–signed assignments document systematic bias: in Germany, students with foreign–sounding names receive lower grades than peers with German names for identical performance (Sprietsma, 2010), a pattern replicated in France and Belgium. On gender, Wang et al. (2018) find that teachers grade girls' language skills and boys' mathematics more favorably even at equal measured achievement — extending to everyday classroom prompts, wait time, and feedback type, as Brophy and Good (1970) had already shown through largely unconscious teacher behaviors.

Bias is most concentrated at institutional transitions, above all guidance decisions at the end of primary school: De Boer et al. (2010) show teacher expectations predict students' secondary track

more strongly than objective, standardized achievement does. This asymmetry is sharper for ethnic–minority and low–income students: in Switzerland, Neuenschwander et al. (2021) find that migration background and socioeconomic status distort teacher expectations in mathematics and German independent of actual achievement — so that guidance tends to reproduce rather than compensate for initial social inequality.

Long–Term Effects

Chetty, Friedman and Rockoff (2014), using U.S. administrative data on millions of students, show that teacher quality — including assessment practices and expectations conveyed — has measurable effects on students' earnings two decades later: school assessment decisions are not isolated episodes but part of biographical trajectories. Developmental research adds that expectations shape not only achievement but students' beliefs about their own ability: Friedrich et al. (2015) find that fifth–graders' mathematics self–concept mediates the link between teacher expectations and grades, a result replicated in Poland by Szumski and Karwowski (2019).

A Final Critical Proposal: Toward a Pedagogy of Suspended Judgment

This evidence supports a theoretical proposal that identifies the conditions sustaining the problem, not just the problem itself. First, the literature has largely treated expectations as individual, correctable psychological failings — the biased teacher — rather than as the system–level effect the inter–teacher variance data suggest: expectations form and stabilize within institutional contexts that structure, legitimize and make them resistant to change. Second, the assessment process is epistemically circular: the accuracy of expectations is validated by the very outcomes those expectations help produce — a circularity that better measurement cannot resolve, because the problem is ontological, not technical.

Third, and perhaps most pedagogically relevant, assessment remains anchored to past achievement: as long as it primarily certifies what a student has already shown, the hidden teacher effect stays structurally invisible — and therefore operative. Vygotsky's zone of proximal development offers a grounded alternative: assessment oriented to potential, rather than current performance, would require teachers to keep open the question of what a student might become rather than closing it around what the student has been — not naive optimism, but an epistemological proposal: the methodical suspension of judgment as a condition for less reproductive assessment.

In sum, the hidden teacher effect is neither an illusion nor an individual deficit: it is a mechanism of social reproduction embedded in ordinary school practice. Making it visible is a necessary, though not sufficient, step toward assessment that does not simply photograph existing inequalities but avoids actively perpetuating them.

Conclusions

This paper takes its cue from the direct observation of a small school setting in which instructional continuity appeared to be associated with more stable, more confident learning, and with a smaller gap between internal grades and external, standardized results. From these elements, together with the extensive literature on the individual dimensions of the "hidden teacher effect," a research hypothesis for investigation on a larger scale has emerged.

The central hypothesis is that instructional continuity in the final three years is associated with a smaller gap between the admission grade (the internal judgment) and the INVALSI score (the external measure), and that this gap retains predictive power for the outcome of the state examination even after controlling for the INVALSI score itself — the same logic that Gentrup and colleagues (2020) applied to teachers' expectations in the early years of schooling is here transposed to the end of the secondary pathway and to continuity as the explanatory variable.

Transferring the hypothesis to a larger scale requires, first, a standardized way of coding continuity from the "15 May" documents (for example, the share of subjects taught by the same teacher across the three–year period, or an index weighted by the hourly weight of the subjects involved) across many schools rather than just one. Multilevel analysis models would need to be adopted, capable of separating individual–level from class–level variance, consistent with the evidence that a substantial

share of the variance in expectations lies precisely at the class level. Another factor to take into account in this type of analysis is the initial socioeconomic level (via the ESCS index already recorded by INVALSI) and the school track, since the stability of the teaching staff varies systematically between academic, technical and vocational schools. One could envisage a pilot study on a purposive sample of institutions within a single Regional School Office, selected so as to vary in track and in observed continuity — relying entirely on documentation already produced by the schools, with no need for new measurement instruments, and with the added advantage that the data do not depend on self–reported evaluation by the teachers themselves.

If the hypothesis were confirmed on a larger scale, instructional continuity would emerge as an additional structural lever for equity, one that does not depend on individual teachers' training and is, in principle, within reach of school–organization and staff–assignment policies, which in most European systems, Italy included, do not currently treat continuity as an explicit goal. If, on the other hand, it were not confirmed once other variables are controlled for, this null result would also be useful, since it would shift attention to other structural levers such as class size, teacher training, and the genuine sharing of assessment criteria.

Either way, a question left open by the international literature could be made observable using data that schools already hold.

A limitation worth stating plainly: continuity is not randomly assigned. Classes perceived as easier to teach, or schools with more stable staffing, may combine greater continuity with better learning conditions for reasons unrelated to the mechanism hypothesized here — a reverse–causality problem no observational design fully resolves. "15 May" documents are also not standardized nationally, so large–scale collection would require real harmonization work. These limits do not undercut the hypothesis but define honestly what testing it would require: what is offered here is a research design awaiting execution, not a conclusion.

More broadly, the hidden teacher effect is not an individual failing correctable with extra attentiveness, but a mechanism rooted in ordinary assessment practice and in the structural constraints of the system. Instructional continuity should therefore be understood not as a definitive fix but as one structural condition — alongside teacher training, transparent assessment criteria, and a plurality of measurement tools — that can reduce, without eliminating, the arbitrariness built into every evaluative judgment. Making it visible remains the necessary, if not sufficient, precondition for a school able to measure learning without silently prescribing it through its own practices of attribution.

Keywords: Hidden teacher effect; Instructional continuity; Teacher expectations; Educational inequality

SESSION 5. EDUCATIONAL POVERTY: CONCEPTUAL AND METHODOLOGICAL APPROACHES THROUGH THE INTEGRATION OF OFFICIAL STATISTICAL SOURCES

ORGANIZER: ISTAT

COORDINATORS: BARBARA BALDAZZI – RINA CAMPORESE

4TH NOVEMBER: 4.30 P.M. – 6.30 P.M. {AULA B0.1 – RESEARCH 6}

Evaluating educational poverty in the metropolitan area of Naples: a hierarchical composite indicator approach

Rosaria Romano – Cristina Davino

Introduction

Educational Poverty (EP) is a *multidimensional latent construct* defined as the deprivation experienced by children and adolescents of the opportunity to learn, experience, develop, and freely cultivate their abilities, talents, and aspirations (Save the Children, 2014). This definition shifts the focus from traditional, outcome–oriented metrics such as early school leaving or dropout rates to the *structural endowments and opportunities* that precede these outcomes. Grounded in Amartya Sen's *Capability Approach* (Sen, 1999), EP is conceptualized as a deprivation of the *substantive freedoms* required to achieve potential developmental trajectories.

This research addresses a critical gap by moving beyond realized achievements to measure the preceding opportunity–based landscape. Adopting an *Ecological Approach* based on Bronfenbrenner's theories (1979), we treat the individual as embedded within nested systems: family, school, and community. These systems collectively shape the student's scope for action. The primary objective is to develop a *Composite Indicator* using *Hierarchical Multiple Factor Analysis* (HMFA) to quantify territorial heterogeneity in the Metropolitan Area of Naples, ensuring a balanced integration of diverse socio–environmental spheres.

Data and survey design

The empirical analysis is based on a stratified probability sample of 3,834 upper secondary students (ages 14–19) across the city and province of Naples. A two–stage sampling design was implemented, using schools as first–stage units and students as second–stage units.

The following table summarises the estimation domains:

Domain Category	Specific Units/Segments
Province of Naples	9 areas (e.g., Phlegraean Area, Sorrentine Coast, Nolana Area)
Municipality of Naples	10 Municipal districts (M1–M10)
School Type	Technical/Vocational Institutes vs. Academic Lyceums
Grade Level	Grades II–III (ages 15–16) vs. Grades IV–V (ages 17–19)

Table 1: Estimation domains in the two–stage sampling design

Measurement of the indicators was conducted via a seven–point Likert scale (1: Strongly disagree to 7: Strongly agree), capturing student perceptions across the hierarchical levels of *educational opportunity* (family, school, community).

Methodology: Hierarchical Multiple Factor Analysis

We employ *Hierarchical Multiple Factor Analysis* (HMFA) (Le Dien & Pagés, 2003; Pagés, 2014), an extension of *Multiple Factor Analysis* (MFA) (Escofier & Pagès, 1994) specifically designed for nested data structures. HMFA is essential for social indicator construction because it prevents blocks with a

high number of variables from dominating the analysis. For instance, without HMFA, the *Family* domain (comprising 20 indicators) would statistically overwhelm the *Context* domain (comprising only 7 indicators), regardless of their relative theoretical importance.
The model utilizes a three–level hierarchy:

- P0 (Highest Order): Global Composite Indicator of Educational Opportunities.
- P1 (Domains): Family, School, and Context.
- P2 (Sub–domains):
    - *Family:* Resources, Cultural/Social Encouragement, and Economic Constraints
    - *School:* Infrastructure and Support Services.
    - *Context:* Availability of Services and Environment/Safety.

The hierarchical structure of the Educational Opportunities model is illustrated in Figure 1, showing the relationships among the global indicator (P0), the three domains (P1), and their corresponding sub–domains (P2).

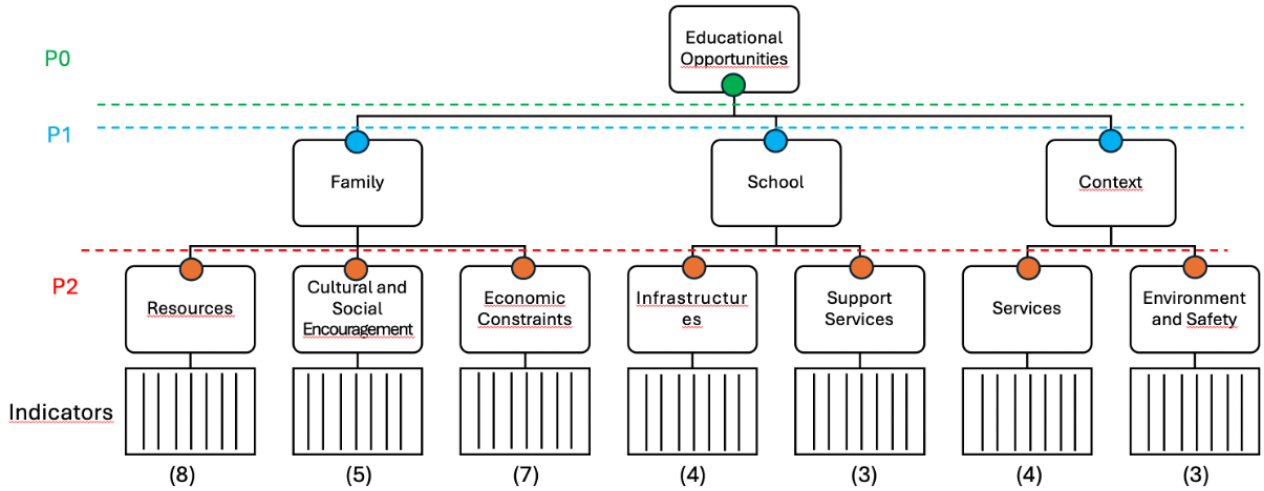


Figure 1: Hierarchical structure of the Educational Opportunities composite indicator

The construction of the composite indicator follows a sequential four–step procedure. First, a Principal Component Analysis (PCA) is applied to each Level 2 matrix (set of indicators) to extract the first eigenvalue, which represents the maximum variance explained within each sub–domain. Next, the original indicators are rescaled by dividing each Level 2 matrix by its corresponding first eigenvalue. This normalization ensures that the maximum variance captured by the first principal component of each sub–domain is equal to one, thereby making the different groups of indicators directly comparable. In the third step, the rescaled indicators are aggregated into the three Level 1 domain matrices, and a new PCA is performed on each domain to obtain the corresponding first eigenvalue. Finally, a second rescaling is carried out by weighting the Level 2 matrices according to the inverse of the first eigenvalue of their corresponding Level 1 domain. The resulting matrices are then combined into a single global matrix, on which a final PCA is performed to derive the overall composite indicator of Educational Opportunities.
HMFA offers several advantages over simple arithmetic weighting:

- **Balanced Group Contributions:** It balances the role of variable groups at each hierarchical level, preventing dominance by large blocks.
- **Multi–level Interpretation:** It provides partial projections, allowing for the analysis of specific levels alongside the global *compromise space*.
- **Theoretical Consistency:** It respects the nested structure of the *Ecological Approach*, maintaining the integrity of sub–domains within their broader domains.

Results and Discussion

The results reveal two primary latent dimensions:

- **Dimension 1 (24.1%–29.4% variance):** Labelled as *School–Based Compensation for Family Disadvantage*, this dimension reflects the interplay between material family resources (study space, digital tools) and school infrastructure. It identifies areas where school quality acts as a critical offset to household deprivation.
- **Dimension 2 (approx. 22.2% variance):** This dimension captures *Community Resource Availability*, focusing on territorial services (cultural facilities, social spaces) and the environmental context (safety and street cleanliness).

Territorial heterogeneity is visualised through the *consensus space*, which reveals how different zones cluster according to their multilevel profiles:

- **Territorial coherence:** Areas like the Vomero (M5) represent affluent clusters with high scores across all domains. Conversely, peripheral districts like Scampia (M8) and Ponticelli (M6) consistently display high levels of deprivation.
- **Partial clouds and domain pull:** The spread of *partial points* indicates territorial coherence. A *tight cluster* of partial points suggests a territory has a consistent profile across family, school, and context. A *widespread* indicates domain–specific pulls; for example, the *Sorrento Coast* exhibits high scores on Dimension 2, pulled specifically by superior contextual opportunities, while the *Nola Area* is pulled downward on Dimension 1 due to specific school–related deficiencies.

The stability of the composite indicator was evaluated using a *leave–one–out* procedure. This sensitivity analysis demonstrated that while the overall dimensions remain robust, specific local perturbations (e.g., removing a specific municipal district) can induce slight coordinate shifts, confirming the sensitivity of the index to the metropolitan area's granular territorial variations.

5. Conclusions and Further Developments

The application of HMFA provides a data–driven yet theoretically anchored measure of EP. By explicitly modelling the hierarchy of *Family*, *School*, and *Context*, this approach offers a balanced perspective on how different spheres of life contribute to educational opportunities. Results represent a diagnostic tool for identifying whether a territory requires interventions focused on *school–based compensation* (to mitigate family resource gaps) or *context–based services* (to improve environmental safety and public infrastructure). Next steps include deeper internal stability assessments and the application of *bootstrap resampling* to evaluate the variability of the composite indicators under sampling fluctuations, ensuring the external validity of the territorial rankings.

Keywords: Educational opportunities; Multidimensional measurement; Multiple factor analysis; Territorial inequalities

Urban Deprivation Areas (ADU) as a Territorial Lens on Educational Poverty: Evidence from Milan

Francesco Chiaradia – Keti Lelo

Introduction

Educational poverty is increasingly understood as a multidimensional phenomenon that concerns not only learning outcomes, but also the unequal opportunities through which children and adolescents can access, use and benefit from educational resources. In 2014, Save the Children Italia defined educational poverty as the deprivation of children's and adolescents' opportunities to learn, experiment, develop and realise their abilities, talents and aspirations (Save the Children Italia,

2014). More recently, Istat has contributed to the national measurement framework through estimates published at the Region × DEGURBA level (Istat, 2026). However, the territorial scale at which indicators are available remains a methodological challenge. Measures estimated for broad territorial domains are essential for national monitoring, but they may not capture the intra–urban distribution of opportunities and constraints that children experience in their everyday environments. This study does not propose an alternative measure of educational poverty. Rather, it examines whether an official intra–urban deprivation geography can reveal territorially differentiated conditions of educational opportunity within a single city.

We examine whether network–based walking accessibility to lower–secondary schools can complement official small–area indicators of socio–economic vulnerability, focusing on Milan's 23 "Aree di Disagio socio–economico in ambito Urbano" (we will call them "Urban Deprivation Areas", or ADU), identified by Istat through the 2021 Index of Socio–economic Disadvantage of Individuals and Families (IDISE). ADU are not administrative areas, but they are contiguous aggregations of census sections constructed through a sequential procedure that aggregates contiguous census sections characterised by high and relatively homogeneous levels of socio–economic disadvantage (Carbonetti et al., 2025; Istat, 2025). This analysis therefore begins from an official sub–municipal geography specifically designed to identify concentrated conditions of economic, labour–market, educational and social vulnerability.

The focus is on the state lower–secondary schools. This level is part of compulsory education and precedes the transition to upper–secondary education, when students and families choose among differentiated school tracks. This transition represents an important educational choice, shaped by previous learning trajectories and socio–economic background (INVALSI, 2025). Focusing on lower–secondary schools therefore allows the analysis to consider a level of compulsory education before accessibility is more strongly shaped by subjective preferences and school–choice decisions. Accessibility is interpreted as a territorially specific condition of educational opportunity: the possibility for youths to reach the nearest state lower–secondary school on foot within a policy–relevant time threshold: the "15–minute city" is used as the theoretical benchmark of pedestrian proximity (Moreno et al., 2021). It is also relevant in Milan, the considered study area, where proximity–based services are part of the municipal policy discourse on "Milano a 15 minuti" (Comune di Milano, 2024).

The study attempts to answer two questions: how is the burden of limited pedestrian access to lower–secondary schools distributed across Milan's 23 ADU and how do different accessibility conditions relate to ADU's socio–economic profiles captured by IDISE and its elementary indicators.

Data used

The study combines official small–area statistics, census–section population data, administrative school data and pedestrian network data. Its starting point is Istat's 2025 release of sub–municipal socio–economic disadvantage data for 2021. This source provides the geometry of Milan's 23 ADU, the Index of Socio–economic Disadvantage of Individuals and Families (IDISE) and its nine elementary indicators of disadvantage (DIS1–DIS9). These can be read in three groups: economic and household vulnerability (DIS1–DIS3), labour–market conditions (DIS4–DIS6) and educational disadvantage (DIS7–DIS9). IDISE values are referenced to the municipal average, set equal to 100: they are therefore comparable across sub–municipal areas within Milan, but not across different municipalities, while its elementary indicators can be compared among different cities. Finally, ADU type is used for descriptive stratification and should not be interpreted as a simple ordinal ranking of severity. The three levels reflect different stages of Istat's identification procedure: level 1 ADU are generally larger aggregations of census sections with medium–high IDISE values, level 2 ADU are smaller areas characterised by higher IDISE values and level 3 ADU are the smallest areas and concentrate the highest IDISE values (Istat, 2025).

Accessibility is measured for the population recorded in the 2021 census variable P16, namely residents aged 10–14 years, used as the small–area demographic proxy for the population potentially

attending lower–secondary school and it does not identify actual school enrolment, attendance or family school choice. The analysis retains the 521 census sections within the 23 ADU with at least one 10–14 years–aged resident, containing 7283 individuals (Istat, 2021). Census–section centroids approximate residential origins. School destinations are state lower–secondary school sites identified in the Ministero dell'Istruzione e del Merito school registry for the 2024/25 school year (MIM, 2025). The analysis combines the 2021 distribution of residents with the 2024/25 school registry and a 2026 network snapshot, calculating walking routes on the OpenStreetMap pedestrian network (Geofabrik GmbH, 2026).

Methodology

Centroids of census sections are used as origins and all state lower–secondary schools as destinations. Walking travel times are computed using "r5r", OpenStreetMap–based pedestrian network data and a walking speed of 4 km/h (Pereira et al., 2021). The analysis adopts a 15–minute threshold and retains a 60–minute maximum routing limit. For each census section, the analysis identifies the walking time to the nearest school and whether the origin lies beyond the 15–minute threshold. ADU indicators are then aggregated using population as a weight. The population–weighted mean travel time therefore represents the average walking time experienced by youth residents within each ADU, rather than the average across census sections. The main indicators are the share of target population living more than 15 minutes from the nearest school, the population–weighted mean travel time and the mean number of schools reachable within 15 minutes.

All 23 ADU are retained in the analysis. IDISE indicators are used to describe the heterogeneous socio–economic and demographic profiles associated with different accessibility conditions. ADU type is used for descriptive stratification and is not interpreted as an ordinal measure of severity. The three ADU types derive from different stages of the Istat construction procedure, which combines progressively stricter disadvantage thresholds with smaller minimum population requirements. They therefore reflect the process through which areas were identified, rather than a simple ranking of severity.

Results and discussion

Figure 1 should be read as a profile of the components of the Milans's ADUs. ADU were identified because they concentrate high overall socio–economic disadvantage as measured by IDISE and variation across DIS1–DIS9 is therefore expected and is analytically informative. In this direction, the educational indicators are particularly relevant: DIS7 measures the share of residents aged 25–64 with low educational attainment, DIS8 the share of young people aged 15–29 who are neither in employment nor education and DIS9 the share of lower– and upper–secondary students who abandon school or do not pass the year.

The figure shows that high accessibility burdens and high educational disadvantage do not coincide in all the situations. ADU 016, a Type 3 ADU, records both the highest IDISE value (111.1) and the highest DIS9 value (26.7), but has no target residents living beyond 15 minutes from the nearest state lower–secondary school. ADU 003 provides a second counter–case: it records a relatively high DIS9 value (16.7) and contains 794 P16 residents, but none of them are located beyond the threshold. Conversely, the most severe accessibility cases do not correspond to the highest DIS9 values: ADU 005 has 92.4% of P16 residents beyond 15 minutes and a DIS9 value of 11.9; ADU 012 has 100.0% beyond the threshold and a DIS9 value of 14.0; ADU 017 has 100.0% beyond the threshold and a DIS9 value of 13.1. Some cases instead combine relatively high accessibility burdens with more pronounced educational disadvantage. ADU 015 records 58.3% of residents beyond 15 minutes and a DIS9 value of 16.3; ADU 013 records 46.5% beyond the threshold and a DIS9 value of 15.4. Their absolute numbers remain limited, however, with 49 and 46 P16 people respectively beyond the threshold.

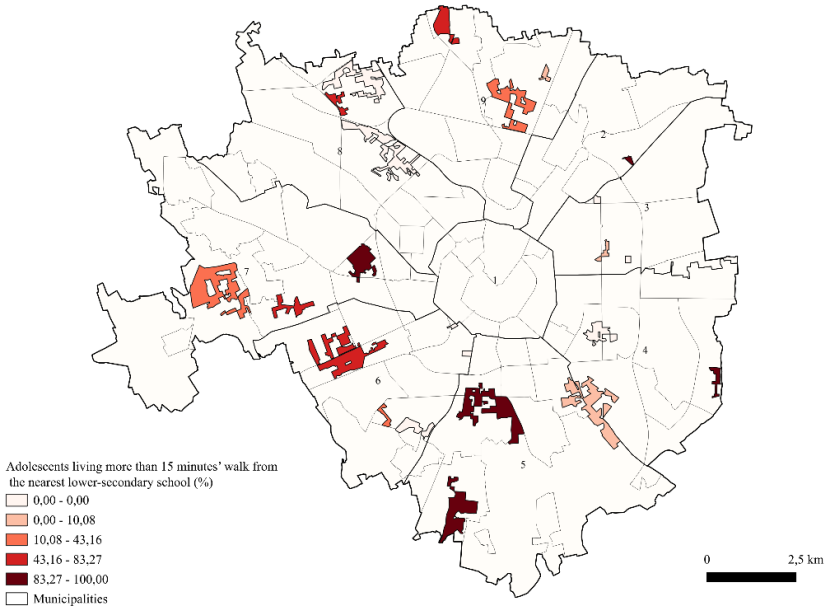
Figure 1. IDISE and elementary indicators on Milan's ADU.

	IDISE	DIS1	DIS2	DIS3	DIS4	DIS5	DIS6	DIS7	DIS8	DIS9
ADU 001 | Type 1	106.6	31.4	14.3	34.5	68.6	19.1	16.1	36.8	31.6	12.1
ADU 002 | Type 1	104.4	23.0	12.1	34.1	72.7	18.0	15.3	33.3	24.3	12.8
ADU 003 | Type 1	106.7	25.5	13.4	34.3	65.9	18.3	17.1	43.5	27.3	16.7
ADU 004 | Type 1	108.4	42.4	16.4	40.2	68.4	18.4	16.9	36.1	35.8	14.5
ADU 005 | Type 1	104.3	24.8	13.9	29.3	74.2	18.0	15.3	29.1	24.3	11.9
ADU 006 | Type 1	105.8	26.6	14.2	34.6	70.2	18.2	16.0	34.5	28.5	14.0
ADU 007 | Type 1	105.8	34.3	14.9	33.1	72.5	17.9	15.8	31.6	26.7	14.9
ADU 008 | Type 1	106.4	28.3	13.2	32.0	64.8	18.4	16.3	40.3	29.3	13.2
ADU 009 | Type 1	104.1	20.3	10.9	26.2	71.9	15.4	15.3	37.1	26.9	11.8
ADU 010 | Type 1	106.5	30.6	15.7	34.1	69.1	19.2	15.4	36.9	27.3	12.4
ADU 011 | Type 2	106.2	28.1	14.2	31.4	67.4	19.5	16.0	39.3	28.4	13.5
ADU 012 | Type 2	106.1	23.7	11.7	34.4	70.1	14.7	18.0	42.2	34.2	14.0
ADU 013 | Type 2	106.3	24.2	14.5	34.3	67.5	20.8	16.7	42.2	23.4	15.4
ADU 014 | Type 2	106.0	23.5	10.6	32.8	65.1	18.0	16.3	41.0	32.1	12.0
ADU 015 | Type 2	108.5	47.8	18.8	39.7	68.3	20.4	17.3	36.8	31.7	16.3
ADU 016 | Type 3	111.1	54.8	20.0	36.9	61.2	27.3	16.0	38.8	29.8	26.7
ADU 017 | Type 3	109.9	46.5	25.2	40.9	62.3	26.1	18.8	40.3	27.5	13.1
ADU 018 | Type 3	109.3	44.3	20.5	36.6	61.2	30.0	17.9	39.7	29.5	9.4
ADU 019 | Type 3	109.0	52.9	23.1	38.7	66.8	25.2	16.0	33.0	23.3	12.5
ADU 020 | Type 3	108.1	31.2	17.1	40.0	67.7	20.6	19.7	40.4	34.9	19.8
ADU 021 | Type 3	109.0	33.3	16.8	38.0	61.1	25.9	18.0	48.0	31.8	16.0
ADU 022 | Type 3	108.9	46.7	14.2	39.3	66.0	22.6	17.0	40.8	37.0	10.0
ADU 023 | Type 3	107.0	37.6	18.4	36.2	70.8	21.6	20.5	30.7	25.1	16.0

Note: cell labels report raw values. Colour shading is scaled independently within each indicator to facilitate comparison across ADUs. For DIS4, which measures the employment rate among residents aged 25–64 years, darker shading indicates lower employment rates and therefore greater disadvantage. Source: Authors' elaboration on Istat 2025.

Across Milan's 23 ADU, 2676 of 7283 10–14 years–old residents, corresponding to 36.7%, live more than 15 minutes' walk from the nearest state lower–secondary school. The population–weighted mean travel time is 13.6 minutes. The 15–minute benchmark is therefore not achieved uniformly across the ADU universe. Figure 2 shows the ADU's geography on Milan's territory.

Figure 2. Share of population aged 10–14 years old living more than 15 minutes' walk from the nearest state lower–secondary school, by ADU.



Source: authors' elaboration on Istat 2025.

Accessibility burdens differ across ADU levels. Type 1 ADU contain the largest absolute number of adolescents located beyond the 15–minute threshold because they are generally more extensive and demographically larger. Type 2 ADU show the highest weighted share of P16 residents beyond the threshold. Type 3 ADU have a much smaller aggregate burden, although some individual cases display extreme percentages because of their small target populations.

Table 1. Accessibility burden by ADU type

ADU level	ADU No.	Residents 10–14 years No.	P16 >15 min No.	P16 (%) >15 min within type	P16 (%) >15 min (considering all)
Type 1	10	6106	2206	36.1%	82.4%
Type 2	5	793	396	49.9%	14.8%
Type 3	8	384	74	19.3%	2.8%
Total	23	7283	2676	36.7%	100.0%

Source: authors' elaboration on Istat 2025, MIM 2025 and OSM 2026.

Among Type 1 ADU 004, ADU 005 and ADU 008 combine very high shares of residents beyond the threshold with large absolute numbers of interested adolescents. ADU 004 includes 669 residents beyond 15 minutes, corresponding to 87.8% of its target population. ADU 008 includes 475 residents beyond the threshold, corresponding to 83.5%, and records the highest population–weighted mean travel time among the large Type 1 cases, at 22.6 minutes. ADU 005 includes 448 residents beyond the threshold, corresponding to 92.4%. ADU 001 has a lower incidence, 49.1%, but still includes 363 affected residents because of its larger P16 population. Type 2 ADU are internally heterogeneous. ADU 012 represents an extreme case, with all 168 P16 residents located beyond the threshold. ADU 011 includes 133 residents beyond the threshold, corresponding to 49.4% of its target population. ADU 013 and ADU 015 show more limited absolute burdens, 46 and 49 residents respectively, but their shares beyond the threshold remain substantial, at 46.5% and 58.3%. Conversely, ADU 014 includes 173 P16 residents but none beyond the threshold. Type 3 ADU require caution because they contain very small target populations: ADU 017 records 100% of its 49 target residents beyond the threshold, while ADU 016 has no one beyond the threshold and a population–weighted mean travel time of 4.8 minutes.

Conclusions

This study proposes a reproducible way to integrate official small–area statistics with network–based walking accessibility in the study of educational opportunities. In Milan's ADU, more than one third of the target population lives beyond a 15–minute walk from the nearest state lower–secondary school. However, this burden takes different forms across ADU levels: Type 1 areas concentrate the largest number of affected adolescents, Type 2 areas show the highest relative incidence and Type 3 areas require caution because small denominators can produce extreme percentages.

The analysis identifies a territorial condition of educational opportunity that can complement official vulnerability indicators. For policy, the approach can flag ADU in which school provision, walking conditions and local mobility should be assessed jointly with socio–economic disadvantage. Decisions on school provision would still require information on capacity, enrolment, route safety, public transport and school choice. By using an official deprivation geography rather than citywide averages, the approach can support a more territorially differentiated identification of areas requiring further assessment.

Keywords: Educational poverty; Accessibility; Spatial inequalities; Small–area statistics

Measuring Educational Poverty at Local Level: Adapting the ISTAT Framework to Tuscan Health and Social Care Zones

Silvia Duranti – Maria Luisa Maitino – Letizia Ravagli

Introduction

Definition of education poverty.Educational poverty can be understood as a multidimensional condition in which children and young people do not have equal opportunities to develop the knowledge, skills and relationships needed to participate fully in education and society. It cannot be captured only through learning outcomes or school attainment, since these outcomes are shaped by the availability and quality of educational resources, services and opportunities in the school and local context. For this reason, measuring educational poverty requires an integrated set of indicators able to account both for educational outcomes and for the conditions that make learning and participation possible.

Literature. Over the last decade, the concept of educational poverty has progressively emerged in both European and Italian debates, although with different emphases. In European policy documents, educational poverty has been mainly defined with reference to minimum educational standards. This approach is consistent with the earlier conceptualisation proposed by Allmendinger (1999) and Allmendinger and Leibfried (2003), who framed educational poverty through two main dimensions: educational attainment, referring to participation in education and formal qualifications, and educational achievement, referring to the acquisition of competences. In particular, according to the European Commission (2014) individuals are considered poorly educated when their level of education remains below upper secondary education or when their competences fall below minimum proficiency levels. This interpretation was further consolidated in the *Education and Training Monitor 2015*, where educational poverty was associated with the share of young people failing to reach minimum standards in education, in terms of both attainment and achievement (European Commission, 2015). In Italy, an early contribution in this field was provided by Checchi (1998), who used the notion of "educational deprivation" to refer to the failure to reach an educational level sufficient to ensure economic autonomy and full social inclusion. This perspective is also connected to the capability approach, since educational disadvantage is not only a matter of years of schooling or qualifications, but also concerns the real opportunities individuals have to develop basic competences and transform them into effective functioning.In the Italian debate, an important step towards a broader operationalisation of educational poverty was made by Save the Children Italia. This approach contributed to framing educational poverty not only as a consequence of poor school performance, but as a condition closely linked to the opportunities available to children and adolescents in their everyday educational environments, emphasising the circular relationship between broader socio–economic disadvantage and educational poverty. Moreover, the concept of educational poverty reflects the innovative shift in poverty research from one–dimensional to multidimensional approaches (Kakwani & Silber, 2007), since it requires attention to several interconnected dimensions of children's lives and educational experiences. The first Educational Poverty Index devoleped by Save the Children Italia (2014), adopted a clearly multidimensional

perspective and was designed to compare Italian regions. It was organised around two main components. The first one focused on educational opportunities within the school system, including early childhood education services, full–time schooling, school meals, school buildings and digital infrastructure. The second extended the analysis beyond school, considering children's and adolescents' access to cultural, recreational and social opportunities, such as theatre, museums, concerts, sport, internet use and book reading. In this first version, educational poverty was therefore understood not only as a lack of school resources, but also as a limitation in the broader set of cultural and experiential opportunities available to children and young people.A second version of the index, developed in 2016, partly revised this structure, while maintaining the regional scale of analysis (Save the Children Italia, 2016). While retaining attention to cultural participation and educational provision, it introduced a more school–oriented measurement strategy. The index was reorganised into two sub–indices: "Learning and Development" and "Educational Provision". The former included school–related outcomes such as low achievement in mathematics and reading and early school leaving; the latter focused on the availability and quality of educational services and school infrastructure. Compared with the 2014 version, the 2016 index therefore placed greater emphasis on school outcomes and educational provision, although cultural participation was still considered within one of the sub–indicators. The work of the ISTAT inter–institutional scientific Commission represents a further institutional step in this trajectory. Rather than relying on a single experimental index, the Commission developed an official statistical framework aimed at defining and measuring educational poverty through a broad set of indicators drawn from multiple official sources. The framework adopts a multidimensional understanding of educational poverty, referring to children and young people aged 0–19 and distinguishing between educational poverty of resources and educational poverty of outcomes.The first domain concerns the educational and cultural resources available in the community of reference, including the family, the school, places of learning and social aggregation, and the wider territorial context. It includes indicators related to family socio–economic and cultural background, housing and material conditions, parental social and cultural participation, the supply and use of educational services, the adequacy of school facilities and services, and the characteristics of the local environment. The latter dimension also captures cultural and social opportunities, such as the presence and use of libraries and museums, green areas, reading habits, sport participation and relationships with peers. The second domain refers to outcomes, understood not only as school achievement in a narrow sense, but as the acquisition of cognitive, personal and social competences. It includes literacy, numeracy and linguistic skills, digital, scientific, financial and civic competences, as well as relational and emotional competences, trust, and regularity of school careers. A distinctive feature of the ISTAT approach is the integration of heterogeneous official data sources. The Commission selected 78 indicators, including 60 indicators of resources and 18 indicators of outcomes, using administrative and survey data sources. This multi–source structure makes it possible to describe educational poverty as a complex phenomenon, in which school outcomes are interpreted together with the conditions and opportunities that contribute to producing them.

Our contribution. The present contribution starts from the ISTAT framework but addresses a different territorial and policy question. The national measurement exercise developed by ISTAT provides a sub–regional reading based on the intersection between regions and the

degree of urbanisation of municipalities. This solution is suitable for national comparison and for identifying broad territorial patterns, but it does not fully capture the local geographies through which regional policies are planned and implemented. It also risks concealing substantial heterogeneity within the same degree of urbanisation, a particularly relevant issue in a region such as Tuscany, where territories classified in similar urbanisation categories may differ considerably in terms of accessibility, demographic trends, school provision and socio–economic conditions.

For this reason, the analysis is carried out at the level of Health and Social Care Zones, 28 aggregations of Tuscan municipalities responsible for the organisation and management of social policies at local level. The choice of this territorial scale is also linked to the broader policy–oriented framework within which the proposed educational poverty indicator is developed. The indicator presented here is conceived as one component of a wider multidimensional measure of poverty, which also includes economic, social and cultural dimensions, further developing a line of work first outlined in Regione Toscana (2024). This territorial scale allows the indicator system to be aligned with a level at which local policies are actually planned, coordinated and implemented. The expected added value is therefore not to replace the national multidimensional framework, but to make it operational for regional and local planning by identifying educational disadvantages at the level at which many interventions can be designed and coordinated.

Data used

The proposed indicator system is based exclusively on official statistical and administrative sources available at sub–regional level. The main sources of data are: SIRIA (the Regional Information System for Early Childhood Services of Tuscany) and Indagine su nidi e servizi integrativi per la prima infanzia (Istat) for early childhood education services data; the open data of the Ministry of Merit and Education (Ministero dell'Istruzione e del Merito, MIM) for the distribution of schools across the territory, class size and school time, as well as for support teachers and the school buildings registry; INVALSI for learning levels; Census data for the reference population of children and young people; and the Tuscan Regional Student Registry (Anagrafe Regionale Studenti, ARS), which provides cross–sectional data on students by municipality of residence, as well as information on grade repetition.

The original territorial units of the data vary according to the source. Some indicators are available at municipal level, others at school or school–complex level, while others refer to individual students or aggregate school populations. The analysis is primarily conducted at the municipal level and subsequently aggregated to Tuscan Health and Social Care Zones. An exception concerns INVALSI data, which are not released at municipal level when fewer than three schools per grade are present; in these cases, the analysis is carried out directly at the level of Health and Social Care Zones.

Methodology

The methodology consists of adapting the ISTAT framework on educational poverty to a more local territorial scale and to the set of indicators that can be reliably calculated for Tuscan Health and Social Care Zones. This implies a selective use of the original framework. The ISTAT framework includes both educational poverty of resources and educational poverty of outcomes, with dimensions referring to family context, school context, territorial, social and cultural context, cognitive competences, and personal and social competences. In this contribution, however, the focus is deliberately restricted to education in a narrower sense. This choice reflects the fact that the proposed educational poverty indicator is

conceived as one component of a broader multidimensional measure of socio–economic poverty, where other economic, social and cultural dimensions are addressed separately.
Therefore, some ISTAT dimensions and sub–dimensions are not included. Within the resources domain, only the school context dimension is retained. The family context is entirely excluded, including socio–economic and cultural status, housing and material conditions, and parents' social and cultural participation. The territorial, social and cultural context is also excluded, including indicators on the local living environment and on children's and adolescents' cultural and social participation. Within the outcomes domain, the sub–dimension on digital, scientific, financial, civic and citizenship competences is not included, and the personal and social competences domain is retained only through indicators on the regularity of school careers, while relational competences, emotional competences and trust are excluded.
Some adaptations are also made within the retained sub–dimensions. The adaptation mainly covers the school context. The sub–dimension on the supply of educational services includes indicators on municipal expenditure and places in early childhood education services, the absence of lower and upper secondary schools in the municipality of residence, schools without canteens, gyms, public transport connections, school bus services, computer rooms or technical classrooms, and a proxy for the completeness of upper secondary educational provision. The sub–dimension on the adequacy of educational services includes indicators on school accessibility, class size, empty or overcrowded classes, student–teacher ratios and support teacher discontinuity. Given the reduced number of retained indicators, the sub–dimension on the use of educational services may be treated jointly with adequacy and includes indicators such as students in mixed classes and students not enrolled in full–time schooling.
Several indicators are recalculated or adapted because the exact ISTAT definition is not available at the desired territorial level. The indicator on accessibility for students with motor disabilities is reconstructed from available school–level variables, whereas the indicators on accessibility for students with deafness and on adapted computer stations for students with disabilities are excluded. The two ISTAT indicators on support teachers are replaced by an indicator constructed by the authors on the share of fixed–term support teachers, while the indicator on the ratio between students with support needs and autonomy and communication assistants is not included.
The outcomes domain is also adapted. INVALSI data are considered for indicators on unconsolidated learning outcomes in grade 5 of primary school, risk of implicit dropout in grade 8, and implicit dropout at the end of upper secondary school, consistently with the ISTAT version. Indicators on school career regularity are constructed using administrative sources where possible: early leaving from education and training is approximated through Census information and ARS data, depending on data availability; non–admission to the following school year in upper secondary education is calculated from ARS via the share of students not admitted among enrolled students; plurirepetition is taken from ARS or, if not available, replaced by a proxy based on the gap between students' age and attended grade, with caution in the interpretation for foreign students. The administrative indicator on upper secondary school abandonment is not retained in the proposed set, as it is considered partly redundant with other indicators of school career irregularity. The indicator on students' self–perceived school ability is also excluded, since it is not available at the required sub–regional scale.

The aggregation strategy will compare different methods for combining the selected indicators, with the aim of assessing which approach provides the most informative and policy–relevant territorial measure. After checking the direction of each indicator, variables will be normalised so that they are expressed on a common scale and can be interpreted in the same direction. Several aggregation strategies will then be tested, including the Adjusted Mazziotta–Pareto Index (AMPI), in line with the ISTAT methodology, and data–driven approaches such as Principal Component Analysis. Different versions of the indicator will also be explored, including more parsimonious specifications, in order to assess whether some sub–indicators can be excluded when they are highly correlated or provide limited additional information.

Results and discussion

The empirical implementation of the indicator system is currently in progress. However, previous IRPET research provides some hypotheses on the expected territorial patterns. First, the resource component is expected to show greater territorial variability than the outcome component, since school provision, accessibility, facilities, class organisation and the availability of educational services are strongly shaped by local demographic density and settlement patterns (IRPET, 2012; IRPET, 2023).

With regard to outcomes, territorial differentiation is expected to be more limited in primary and lower secondary education, where previous analyses have shown a relatively homogeneous regional pattern, and more pronounced in upper secondary education (IRPET, 2021). Earlier research on Tuscany has shown that educational fragility tends to accumulate over the school career through low achievement, grade repetition, irregular trajectories and school failure, becoming more visible in the transition to and within upper secondary education (IRPET, 2014). For these reasons, the expected results are unlikely to reproduce a simple "centre versus periphery" pattern. Rather, the indicator system is expected to identify differentiated territorial profiles and internal heterogeneity across sub–dimensions. This is particularly relevant in Tuscany, where territories with the same degree of urbanisation may differ substantially in terms of school provision, accessibility, demographic dynamics and local educational opportunities. Urban areas may show criticalities related to demand pressure and overcrowding, while inner and peripheral areas may be more affected by accessibility constraints, weaker full–time provision, small–scale school organisation and limited variety of upper secondary programmes (IRPET, 2012; IRPET, 2023). At the same time, small class size in peripheral areas may also represent a potential advantage, suggesting that territorial differences in educational poverty cannot be fully interpreted through broad urbanisation categories alone. Recent IRPET work on inner areas suggests that differences may be substantial not only between central and peripheral territories, but also among peripheral areas themselves, depending on accessibility, demographic trends, local school provision and the degree of connection with urban centres (IRPET, 2023).

Keywords: Educational poverty; Territorial indicators; Sub–regional measurement; Local policy

Educational Poverty: The Work of the Istat Inter–institutional Scientific Commission

Barbara Baldazzi – Rina Camporese – Miria Savioli – Elisabetta Segre

Introduction

Educational poverty has attracted growing attention in Italy across academic, political, and institutional spheres, yet the field has long been characterised by definitional fragmentation and measurement heterogeneity. Early contributions proposed a range of different frameworks: Checchi (1998) approached it through the lens of educational returns; Allmendinger and Leibfried (2003) situated it within comparative welfare state analysis; Lohmann and Ferger (2014) formalised the distinction between absolute and relative educational poverty; Glaesser (2022) examined its conceptual and empirical implications. In the Italian context, Save the Children (2014, 2018) pioneered empirical measurement with composite indices focused on children's access to educational and cultural opportunities, while Pratesi et al. (2021) applied small area estimation techniques to produce sub–regional estimates of a multidimensional index. Against this background, Istat established an Inter–institutional Scientific Commission in March 2023 — comprising over 50 experts from academia, international organisations (UNICEF, World Bank, UNESCO), civil society (Save the Children, Fondazione Con i Bambini), and public institutions (INVALSI, Ministry of Education, INPS, Bank of Italy) — with the mandate to produce a shared definition, a conceptual framework, and a statistical measurement system for educational poverty in Italy.

The Commission analyzed national and international literature on the subject, proposed a definition of educational poverty along with a measurement framework, identified existing information sources within the National Statistical System, and selected a broad set of multi–source indicators. It then calculated these values at a sub–regional level—by crossing administrative regions with the municipalities' urbanization levels—using small–area estimation methods for sample–based sources. Finally, to synthesize the information and outline priority areas for intervention, the Commission constructed several composite indices and mapped the national territory, thereby laying the groundwork for regular, up–to–date measurement of this phenomenon.

The Commission concluded its work in January 2026. The main results have been presented at the conference *"Measuring Educational Poverty: Results and Perspectives of the Work of the Inter–institutional Scientific Commission,"* organized in Naples on April 16 and 17, 2026.

This paper illustrates the Commission's work from a methodological perspective, with a fundamental research question in mind: how to move from a rich but complex multi–source indicator system to actionable, policy–relevant knowledge about a latent, multidimensional phenomenon. The central question is not only how to measure educational poverty, but how to synthesise heterogeneous statistical information in a way that supports territorial prioritisation and informs policy design. An early exploratory analysis underpinning the Commission's methodological choices was presented in Segre, Savioli and Quondamstefano (2025), Pratesi et al. (2025).

The Commission's Work: The Measurement Approach

The Commission's work was guided by a multidimensional approach centered on the broader Italian meaning of "educazione" compared with the English term "education." This expands the focus to a plurality of aspects that go beyond the mere failure to achieve schooling goals, which is instead typical of the Anglo–Saxon concept of "educational poverty." Indeed, an adequate or minimum level of social inclusion in different spheres of life requires not only knowledge and disciplinary skills (schooling/*istruzione*), but also the ability to relate to others and society (education/*educazione*), as well as behaviors, habits, and attitudes directed toward socially shared goals (training and personal development/*formazione*).

Although this issue is relevant at every stage of life, the Commission focused its attention on childhood and adolescence (ages 0–19), the phases in which individuals build the core of their human

and social capital. Furthermore, the Commission recognized the role of the "educating community" (family, school, associations, and other actors actively contributing to the development and growth of children and youth) and the opportunities it provides.

The Definition of Educational Poverty

Educational poverty is a latent variable, a theoretical construct that cannot be directly observed but is instead inferred through measurable indicators. In line with the multidimensional approach, its effects manifest across various dimensions: in cognitive skills, but also in emotional and relational growth.

The Commission recognizes access to educational opportunities as a *condicio sine qua non* for the development of skills and personal growth. Based on this reflection, the Commission adopted a vocabulary promoted by the OECD (OECD, 2021) and commonly used in well–being measurement frameworks. This vocabulary distinguishes individual outcomes (*esiti*) from the resources offered in different life contexts (familial, social, cultural, and environmental) (*drivers*/resources) which, with varying degrees of intensity, affect the opportunities to achieve desired or desirable outcomes.

The Commission defined educational poverty as a complex multidimensional phenomenon—resulting from the familial, economic, and social context in which children and youth (ages 0–19) live—distinguishing it into **poverty of resources** and **poverty of outcomes**:

- **Poverty of resources** is a condition resulting from a lack of educational and cultural resources in the broader reference community (family, school, places of learning and socialization, living environments) or from a limitation in the opportunities for personal growth experiences that such resources provide.
- **Poverty of outcomes** means failing to acquire the personal, social, and cognitive skills necessary for growth and the development of relationships with others, for cultivating one's talents and realizing one's aspirations, and at a collective level, for feeling part of a community, consciously exercising the right to active citizenship, and contributing positively to the country's well–being.

Children and youth growing up in contexts deprived of educational and cultural resources and opportunities are considered **at risk of educational poverty**.

Children and youth who, with specific reference to their age group and achieved outcomes, have not reached the level of education and skills necessary to avoid forms of social exclusion in various spheres of life are considered to be **in conditions of educational poverty**.

With this definition, the Commission distinguishes between the *condition* of educational poverty (which manifests at an individual level in terms of outcomes) and the *risk* of educational poverty (linked to the scarcity of resources and opportunities). This scarcity can contribute—hence the term risk—to creating social, territorial, or individual situations where the right to full development may be compromised.

The Conceptual Framework: Resources and Outcomes

In light of this definition, the Commission developed a conceptual framework functional to the quantitative measurement of the phenomenon, divided into two domains: the **resources domain** and the **outcomes domain**. Each of these two domains is further broken down into a series of dimensions and sub–dimensions that account for the multiple facets of the phenomenon.

The resources domain is divided into three dimensions: the family context, the school context, and the territorial, social, and cultural context in which children and youth live. As for outcomes, two dimensions were identified: cognitive skills on one hand, and personal, relational, and social skills on the other, which map back to the concepts of hard and soft skills typical of economic literature.

	Domains	Dimensions	Sub dimensions
Educational poverty	Resources	Family Context	Socio-economic and cultural status Housing and material goods Parents' relationships, social and cultural participation
		School Context	Provision of educational services Adequacy of educational services Use of educational services
		Territorial, Social, and Cultural Context	Living environment Children and youth's relationships, social and cultural participation
	Outcomes	Cognitive Skills (Hard Skills)	Alphabetical, numerical, linguistic Digital, scientific, financial, civic and citizenship
		Personal, Relational, and Social Skills (Soft Skills)	Relational Emotional Trusting interaction Regularity of the school career / educational path

Once the conceptual reference framework was defined, the Commission launched a process to define and select statistical indicators to measure its various aspects. This process took into account both the chosen level of territorial detail for calculating the indicators (the degree of geographical disaggregation) and the decision to calculate deprivation indicators.

Methods: The Territorial Level

The Commission opted for a territorial mapping that crosses the regional level with the municipality's degree of urbanization. The choice fell on the **Degree of Urbanisation (DEGURBA)** classification, developed by Eurostat in collaboration with the OECD to distinguish urban, intermediate, and rural areas in a cross–country comparable way. This choice makes it possible to provide a picture of the national territory that looks inside the borders of administrative regions, distinguishing three territorial areas within each region: the first groups together Cities (densely populated areas), the second Towns and suburbs (intermediate density areas), and the third Rural areas (thinly populated areas).

This process identified 62 territorial units of analysis (19 regions and 2 Autonomous Provinces, each disaggregated by the 3 degrees of urbanization, with the exception of Valle d'Aosta, which has no municipalities classified as cities or densely populated areas).

Methods: Indicator Calculation Criteria

The indicators selected by the Commission to measure educational poverty were designed and calculated as **deprivation indicators**, aiming to provide a clear picture of hardships regarding resources and outcomes directly from the individual indicators within the data matrix.

To calculate these indicators, administrative sources were used—such as the Ministry of Education and Merit (MIM), INVALSI, and SIAE—alongside Istat census surveys (Survey on the inclusion of students with disabilities, Survey on early childhood educational services, Survey on libraries, and Survey on museums and similar institutions), the Statistical Register of Non–profit Institutions, and five Istat sample surveys:

1. Multipurpose Survey on Aspects of Daily Life (*AVQ*)
2. Survey on Living Conditions (*Eu–SILC*)
3. Labour Force Survey (*FdL*)
4. Citizens and Leisure Time Survey (*CTL*)
5. Children and Youth: Behaviors, Attitudes, and Future Plans Survey (*B&R*)

Methods: Small Area Estimation

For indicators derived from sample sources—where the region–by–DEGURBA cross–section is not a planned estimation domain—Istat calculated direct estimates for the desired territorial intersection.

The DAGUM Centre, a member institution of the Commission, then applied small area estimation methods to improve their accuracy using the Fay and Harriot method (1979).

Methods: Indicators and Composite Indices

To synthesize the information related to the 78 educational poverty indicators selected by the Istat Commission and to map the national territory, several composite indices were constructed using the Adjusted Mazziotta–Pareto Index (AMPI) method (Mazziotta and Pareto, 2016 and 2024).

The AMPI method was applied to 62 units of analysis, and seven composite indices were calculated: one for the "outcomes" domain, one for the "resources" domain, and five intermediate indices for the dimensions of family, school, territorial, social, and cultural contexts, as well as cognitive, personal, and social competencies.

The indices range approximately between 70 and 130, with a base value of 100 representing the Italian average. The value of 100 can be interpreted as a relative threshold for educational poverty. Areas with values above 100 are in a situation worse than the national average, while units with values below 100 are in a better situation.

This approach provides a snapshot of the phenomenon, achieving a twofold objective:

1. Identifying the territorial areas where the risk and presence of educational poverty are highest.
2. Offering an information framework to guide public investments and policies.

Results and discussion

An analysis of the distribution of the composite index of educational poverty risk, based on resources, reveals more varied patterns by degree of urbanization. Sicily, Calabria, and Campania face the most difficult situation across their territories, although the intensity varies by municipality type. In particular, in Sicily, while all territories exhibit a resource shortage exceeding the national average, rural areas show much higher levels of deprivation than both suburbs and, above all, cities.

Among the regions of Southern Italy, the cities and suburbs of Sardinia, Abruzzo, Basilicata, and Molise, as well as those of Abruzzo, exhibit a range of values that are relatively less negative (around the national average). In the other regions as well, the degree of urbanization has a significant impact on the index's performance. In the Central–Northern regions, the situation is generally better; in particular, the risk of resource poverty in cities is lower than the national average. The cities of Trento and Bolzano/Bozen stand out positively, as do cities in the Marche, Tuscany, Emilia–Romagna, and Friuli–Venezia Giulia. Even in these areas of the country, however, inequalities exist depending on the degree of urbanization. Particular attention should be paid to the situation in rural areas of the Northwest and Northeast, especially in Liguria, Veneto, Piedmont, and Lombardy. Compared with findings on the extent of resources, the composite educational poverty index based on outcomes reveals a higher level of inequality across the country as a whole (composite index values range from 91.0 to 125.8). An analysis of the geographical distribution of these values reveals diametrically opposed situations. On the one hand, Umbria, Molise, and Veneto, where all municipalities—regardless of their degree of urbanization—show a clear advantage, with index values lower than the national average. Favorable situations are also observed in the cities of the Marche and Abruzzo regions, in the rural areas of the Aosta Valley and the Autonomous Province of Trento, and in the small towns of the Autonomous Province of Bolzano/Bozen. On the other hand, Sardinia shows a significant disadvantage across all municipalities, regardless of urbanization level, with a composite index value of 125.8 in rural areas. Sicily, Calabria, and Campania also face difficult conditions across their entire territories, with some variations depending on the degree of urbanization in each municipality. In Sicily, the situation is problematic in all municipalities; in Calabria, rural areas have the highest index values, while in Campania, it is the cities. Rural areas of Puglia also experience more difficult conditions.

Among the regions with values worse than the national average are Liguria and Emilia–Romagna. In Liguria, in particular, this is the case in rural areas and small towns, while in Emilia–Romagna it is the case in the mountainous and rural areas of the Apennines.

Conclusions

The Commission's work demonstrates that a carefully designed multi–source indicator dashboard, combined with Fay–Herriot small area estimation and AMPI composite index methodology, can generate policy–relevant territorial knowledge on a complex latent phenomenon.

The resources/outcomes dual architecture produces operationally distinct territorial maps that point to qualitatively different types of policy intervention. Areas characterised by high resource poverty call for supply–side investments in educational infrastructure and services. Areas where outcome deprivation is disproportionate relative to available resources — as in the case of Sardinia — call instead for demand–side or community–level interventions targeting the social and family conditions that mediate the conversion of educational opportunities into individual competencies.

The DEGURBA classification proves analytically productive: by crossing regional boundaries with the degree of urbanisation, it exposes substantial intra–regional heterogeneity that would be invisible at the regional level alone.The composite indices reveal a picture of educational poverty in Italy that is more nuanced and territorially differentiated than the conventional North/South divide would suggest.

Keywords: Educational poverty; Multidimensional approach; Official statistics; Territorial inequalities

Inequalities in Education: An Analysis Based on Well–Being Indicators

Barbara Baldazzi – Lorenzo Di Biagio

Introduction

Education is extremely important for well–being—not only because it is a means of self–actualization, but also because it is one of the main determinants of well–being, both at the individual level and for the progress of society.

Those with higher levels of education typically enjoy a better standard of living, have a longer life expectancy, adopt healthier lifestyles, and have easier access to stable jobs (Baldazzi, 2023).

At the same time, Italy is a country that has historically been deeply diverse. While a process of convergence developed up until the 1970s, this process—particularly with regard to GDP—has now come to a halt.

In this paper, we would like to go beyond GDP, using the Istat–BES framework on education, to investigate the extent and characteristics of regional inequalities in education.

Data Used: The BES Project

The Istat project on Equitable and Sustainable Well–being (Bes) was launched in late 2010 as a joint initiative of the CNEL and Istat, through a dialogue with civil society and academia, as part of the "Beyond GDP" debate (Bacchini et al, 2021).

The objective is twofold: to assess societal progress by integrating economic, social, and environmental dimensions; and to provide policymakers and the general public with a shared measurement framework.

The BES framework comprises 12 domains and 152 indicators. Education and training (and cultural participation) is one of these domains (Istat, 2025).

We will focus our analysis on a selection of high–quality statistical indicators on formal education and non–formal learning from the Bes framework. The indicators are divided into 5 groups.

In the first group (access to early childhood education), we see a steady increase in the percentage of children enrolled in preschool, partly due to the decline in birth rates, which is narrowing the gap between potential users and available spots (European target: 45% by 2030); by the age of 4–5, nearly all children are enrolled in educational programs (target of 96% by 2030), but the percentage had declined (and is now recovering).
In the second group (educational pathways), all indicators have improved over time. The rate of early school leavers has fallen below the European target of 9% for 2030. The number of NEETs is also falling sharply. The indicator for university enrollment, however, shows a modest increase.
In the third group (educational attainment indicators), all indicators are on the rise, but a comparison with other European countries still places Italy near the bottom in terms of high school and university graduates.
In the fourth group (skills), the competence in Italian and Maths is worsening, while digital skills have increased, but are still far from the European target of 80% by 2030.
The fifth group (participation in lifelong learning) comprises a single indicator, which has experienced accelerated growth since COVID–19 driven by online training.

Methods

We considered methods that are sufficiently general and simple to be used within the framework of official statistics and for groups of indicators with different definitions, units of measurement, and sources.
The first method that we examined is standardized differences, i.e., reducing indicators to a common scale in terms of mean and variance, in order to compare different regions and different indicators. This approach is analogous to z–scores but centers the values on the national mean (to simplify interpretation) and reverses the sign if the indicator has a negative polarity (so that the values can be interpreted in terms of well–being: the higher the values, the higher the well–being).
Using the standardized differences method, we effectively filter out the intrinsic variability of the phenomena, thereby identifying the regions that deviate the most from the territorial variability (within the Italian context), pinpointing both positive and negative outliers (which warrant further investigation), as well as regions exhibiting more standard patterns.
The second method is the computation of the coefficient of variation (CV). This index is very useful and is widely used both in the literature and in official statistics to monitor the intertemporal convergence of regions (the so–called σ–convergence.) However, it is not always suitable for comparing disparities between different indicators: even if the CV neutralizes the effect of the unit of measurement and puts the order of magnitude into perspective, it does not account for the intrinsic variability of the phenomenon, and above all, it is influenced by how the indicator is structured. Therefore, it does not ultimately allow us to distinguish between situations that are more or less critical or more or less manageable with regards to disparity.
The third method that we considered is the computation of the percentage of explained variability (eta squared), in our case by the "region" factor. The eta squared measures the percentage of variability attributable to regional classification, relative to the overall variability (Richardson, 2011).

Results and discussion

For most indicators, the South lags behind to a greater or lesser extent, and in general, there is a marked North–South territorial gradient that is particularly evident for NEETs and basic digital skills. There are some exceptions: enrollment in the school system among 4– to 5–year–olds, where the southern regions perform better (almost all above the 2030 European target of 96%); the attainment of STEM degrees; and, to some extent, the transition to university and early exit from the education system (even if the islands perform particularly poorly).
Analyzing standardized differences, we see that the areas of excellence account for less than a quarter of the critical issues, and the average score for the critical issues is, in absolute terms, higher than that for the areas of excellence (–2.1 versus 1.8). When examining the distribution of deviations, a

clear asymmetry emerges: extreme "shortcomings" are significantly more numerous than instances of "excellence," even though nearly 60% of the values are equal to or exceed the Italian average. This data helps us better understand the structural nature of regional disparities in education. On the one hand, we do not find individual regions capable of standing out exceptionally on the positive side (top performers); rather, there are groups of territories—especially in the Center–North—that have medium–to–high values, driving up the Italian average. However, for each indicator, these are fairly compact groups, with no regions standing out positively; Conversely, the tail of negative values points to areas that are indeed severely lagging behind, suffering from a systemic disadvantage that isolates them from the rest of the country (primarily regions in the South, particularly Sicily).

We note that Lombardy, the Autonomous Province of Trento, Emilia–Romagna, Umbria, and the Marche (regions in the North and Central Italy) have values equal to or above the Italian average for nearly all indicators (10 or 11 indicators); in contrast, for Campania, Puglia, and Sicily (all regions in the South), at most two indicators have values equal to or higher than the Italian average. In fact, the regions of Southern Italy generally have at most 4 out of 12 indicators equal to or higher than the Italian average, whereas the regions of Northern and Central Italy generally have values equal to or higher than the Italian average for at least 9 indicators. Positive exceptions in Southern Italy are Abruzzo and Molise, which have profiles comparable to those of some regions in Northern and Central Italy. Negative exceptions in the North and Central regions include Valle d'Aosta, Lazio, and especially Liguria and the Autonomous Province of Bolzano. For the Autonomous Province of Bolzano, one must take into account the sources (MIM), which may introduce distortions; however, even without considering transition to university and STEM degrees, the province still falls below the national average (and compared to the Autonomous Province of Trento) in areas such as college graduation, literacy and numeracy skills, and participation in early childhood education. Liguria does indeed have a negative median but relatively low variability, with one–third of the indicators concentrated around zero; therefore, we can say that it is in a state of "stable mediocrity," similar to the Italian average. Piedmont (ranking first), Tuscany, and the Marche also have lower variability, indicating an education and training system that is homogeneous in its various aspects, but with positive medians. The regions with the highest medians are Umbria, the Autonomous Province of Trento, and Friuli–Venezia Giulia.

Shifting our perspective, we no longer look at the performance of individual regions, but rather at the indicator as a whole to quantify the "weight of geography". This, too, has political relevance: if regional differences have a major impact on indicator levels, then it likely means there are structural differences between regions—in education and training systems, or even simply in resources and infrastructure—that should be investigated and addressed by policymakers. What interests us most is the comparison between the 12 indicators in terms of regional effect and—where possible— in terms of the gender factor. We see that regions are a more significant factor for numeracy skills, NEETs, children enrolled in daycare, and digital skills, while the gender factor is almost negligible. Geography plays an important role in terms of opportunities, the labor market, and infrastructure. Note, however, that for literacy skills, the geographic factor accounts for about one–third of what it accounts for in numeracy skills. Some indicators, on the other hand, are more closely linked to cultural expectations or specific gender biases, where the role of region is less significant and the role of gender is comparable to or even far greater than that of region: both gender and region play a fairly important role in the transition to college, while in the percentage of college graduates, the role of gender is predominant.

Conclusions

Even with regard to education, as measured by the indicators of the Istat–BES framework, there is a marked geographic polarization along the North–South axis, with some exceptions both for certain indicators (e.g., school enrollment of 4– to 5–year–olds) and for some regions that buck the trend relative to their respective group (Bolzano, Liguria; Abruzzo and Molise).

There are few instances of absolute excellence, with large positive deviations; in general there is a fairly compact group of regions—primarily in the center–north—with medium–to–high levels (above the Italian average) regarding education indicators; conversely, there are marked weaknesses in some regions, particularly in the South, and especially in Sicily.
The effect of geographic location on inequality is relatively modest, but it is nonetheless significantly greater than the gender factor and cannot be ignored when it comes to skills, NEETs, participation in early childhood education, and high school graduation rates.

Keywords: Inqualities; Variability; Differences; Regional effects

SESSION 6. DATA LITERACY AND DATA USE IN SCHOOLS: EMERGING PRACTICES, CHALLENGES, AND PROFESSIONAL CULTURES

ORGANIZERS: FONDAZIONE BRUNO KESSLER – FONDAZIONE PER LA SCUOLA
COORDINATORS: DAVIDE AZZOLINI – VERONICA MOBILIO
5TH NOVEMBER: 8.30 A.M. – 10.30 A.M. {AULA MAGNA – RESEARCH 7}

Investigating Educational Data Literacy Needs in Italian Schools: A Multidimensional, Role–Differentiated Approach from the SCHOOLWISE Needs Assessment

Flavio Manganello – Giancarlo Masi – Martina Ragusa – Chiara Fante

Introduction

Educational data literacy figures in European policy as a competence schools are expected to develop, yet its conceptual foundations remain unsettled. A recent comparative review of four research traditions reports that no single model encompasses all five domains of practice (technical, pedagogical, organizational, ethical, and transversal), that the competencies ascribed to teachers and to school leaders are seldom differentiated, and that the organizational domain receives least attention (Manganello et al., in press). The traditions compared are partial in instructive respects: the data–based decision–making and teacher data literacy strands situate competence primarily in the individual teacher and orient it toward school improvement (Mandinach & Gummer, 2016; Schildkamp, 2019), whereas the critical strand maintains that data practices are not neutral but encode assumptions and norms that reflect established regimes of power and knowledge (Raffaghelli & Stewart, 2020). What remains underspecified is how data literacy is configured beyond the individual role, across the scales at which schooling is governed, and where the organizational and infrastructural conditions these frameworks least articulate reside.

The present contribution develops this diagnosis empirically within SCHOOLWISE, an Erasmus+ Forward–Looking project concerned with how schools understand, govern, and use data for improvement. With reference to the Italian strand of the data collection, it addresses three research questions. First, how are data literacy needs configured across professional roles and across scales of action, from the classroom to the national evaluation body? Second, does the role differentiation left underspecified by current frameworks operate as a meaningful distinction when actors describe their own work? Third, at which points do the data that each actor regards as most relevant to its decisions and the infrastructure available to it diverge? They are examined through a stratified qualitative needs assessment that juxtaposes actors at four scales, to render observable the organizational dimension that surveys of individual competence overlook.

Data used

The study draws on two sources of evidence collected in the first half of 2026. The first comprises two online surveys, for teachers and for school leaders and administrative directors (DSGA), mapping self–reported practices, attitudes, and perceived competence across the five domains. The second, on which the present analysis relies, comprises five focus groups and interviews stratified by scale of action across four levels: the classroom teacher, the school leader, the regional school administration (USR), and the national evaluation body (INVALSI). Eleven participants took part: three teachers and the head and deputy head of a comprehensive school (istituto comprensivo) in northern Italy, an associate partner; three officers from two regional school offices; one national–body researcher; and

the head and an Erasmus coordinator of an upper–secondary institute (IIS) in southern Italy. The classroom and part of the institutional level were examined within that northern school, so that the teacher and leadership sessions could be read against one another. Participation was voluntary under the GDPR, with CNR–ITD as data controller and transcripts pseudonymized prior to analysis. The qualitative sessions are complete; the survey corpus is still being assembled, so the findings below are preliminary.

Methodology

The investigation adopts a framework–guided qualitative design, developed by Graz University of Technology (TU Graz) as leader of the project's Work Package 2, and informed by the conceptual framework elaborated by CNR–ITD (Manganello et al., in press). The five–domain matrix functions as an a priori coding structure, applied deductively while remaining open to categories it does not anticipate, with a view to later reading across a novice–to–expert continuum (Beck & Nunnaley, 2021). Semi–structured protocols are differentiated by role, from system use, obstacles, trust, and privacy for teachers, to governance and responsibility for leaders, and to regulation, ethics, and AI at the regional and national levels. All instruments share one operational definition of educational data use, the practical and responsible use of information already embedded in school life (assessment, attendance, observation, progress), as distinct from data as mere figures or instruments of control; this follows the understanding of data literacy as the transformation of information into actionable instructional knowledge (Mandinach & Gummer, 2016) and the school–improvement orientation of data–based decision–making (Schildkamp, 2019). Coding proceeds transversally by role, to establish how a given competence operates differently according to the actor's scope and authority.

Results and discussion

At the classroom level, the teacher focus group, conducted with staff of the northern comprehensive school across primary and lower–secondary grades, yields three recurrent patterns. First, much of the data on which teachers rely to monitor a pupil lies outside the institutional systems: participants maintain individual paper records, consulted to determine whether a child has progressed or regressed, because the electronic register accommodates no observational information of this kind; the special–needs referent reports no institutional format for such a record. Second, the systems in use are experienced as inflexible; one teacher was excluded from a registration platform after exceeding a fixed window, necessitating a manual correction through the office. Third, the teachers articulated a need to systematize informal observation without added workload and to receive it in visual form, which in the discussion took the form of a spoken note organized over time into a documented record. The ethical dimension was prominent yet externalized: privacy recurred as a source of anxiety, sustained by contradictory guidance from above, and was treated as a rule to comply with rather than as professional competence.

At the institutional level, examined in two schools situated in different territories, the pattern is consistent. In the northern comprehensive school, the head and deputy head articulate what the teacher accounts leave implicit. They describe a long–standing wariness toward data and an apprehension about the comparisons it invites, since a measure registering a difficulty may read as a verdict on one's work, attributing that difficulty to the organization and its culture rather than any single instrument. The data returned from higher levels are too coarse for a particular school, since a city–wide unemployment figure conveys little where neighbourhoods differ markedly. Governance rests on a vertical structure in which ultimate responsibility resides with the head whatever the head knows, on a school of weak ties lacking a defined intermediate tier, and on platform choices shaped by cost, by the free provision the pandemic made decisive, and by informal recommendation among

colleagues, with no technical office competent to evaluate the systems adopted, a single technical assistant covering six institutes one day per week.

The second case, an upper–secondary institute in southern Italy, both corroborates and extends them. Its head emphasizes interoperability: identical data must be entered into several national and regional platforms that do not communicate, duplicating the work of secretariat and teachers, while dependence on one management suite renders change prohibitive. She states the responsibility problem plainly: she is accountable for the data she knows and for those she does not, since sensitive records pass through secretariat, teachers, and auxiliary staff. Bureaucratization is experienced as a proliferation of often contradictory directives in which form displaces substance and compliance encroaches upon instruction; the national results, valuable as a starting point for the improvement plan, stand in tension with the individualization they are meant to inform, and assessment transparency conflicts with data protection when pupils ask how they stand relative to peers.

At the regional level, the school–office (USR) referents of two regions describe markedly different arrangements. The first operates a proprietary system, built and maintained by its own staff, issuing structured forms under credentialed access with the head's validation incorporated; the second relies on an external regional company that supports schools and serves as a repository for digital projects, so that part of the data flow passes outside the office. Neither sits within a shared national design, and their data overlap only marginally with those gathered nationally, contact resting on personal acquaintance rather than structure. The referents dwell on responsibility: the head is data controller yet often does not apprehend the role's full extent, while the external data protection officer advises under contract without assuming it. They identify the regulatory framework, from data minimization and the EU–US Data Privacy Framework to the national law (Law 132/2025) adapting domestic legislation to the AI Act, and they characterize student wellbeing and sensitive records as a growing source of friction with families.

At the national level, the evaluation body (INVALSI) describes a structured architecture in which internal and external evaluation proceed through dedicated platforms, returning standardized results and indicators that let each school compare itself with schools of comparable socio–economic composition, down to the class, in the absence of a regional tier, since the body operates nationally. The account does not merely stand above the others but reproduces their tension. The architecture functions but is described as prone to bureaucratization: secretariats enter the input data without apprehending its purpose, so records remain incomplete and indicators fail to populate, and a once co–constructed process becomes, at scale, a matter of compliance. The privacy–pedagogy tension likewise recurs: anonymous longitudinal codes may re–identify individuals once class is cross–tabulated with disability or origin, while support teachers, excluded from the register, are deprived of the data their role requires. A pervasive domestic electronic–register industry, reinforced by pandemic–era reliance on global vendors, has shaped a rigid, summative conception of assessment at variance with accredited pedagogy.

Across the four scales, the accounts address the research questions in ways the frameworks do not anticipate. First, data literacy needs are configured by scale of action no less than by role: a given competence, whether interpreting comparative data, governing a platform, or holding responsibility for a dataset, assumes a different form and meets a different obstacle at each level, so the two–role distinction the frameworks leave underspecified is better understood as scale–differentiated, a reading the two institutional cases, in contrasting territories and school orders, independently corroborate. Second, a structural pattern recurs that description alone does not disclose: at every scale, the data an actor regards as most relevant to its decisions are those least supported by its available infrastructure, from teachers' observational records to leaders' insufficiently granular

returns, regional systems in parallel, and national aggregates stopping at the class. This pattern is here termed an infrastructural mismatch, located in the organizational dimension the frameworks least articulate and that the review reconceives as governance as institutional infrastructure (Manganello et al., in press). Third, responsibility is decoupled from capacity: formal accountability is assigned to actors, and to the head as data controller in particular, who may command neither the competence nor the infrastructure the role presupposes, which lends empirical substance to the contention that data practices are not neutral but reinforce configurations of power (Raffaghelli & Stewart, 2020).

These patterns bear directly upon teaching and learning. The data most pertinent to instruction are formative and observational, yet such data are precisely what institutions are least equipped to retain and aggregate, consistent with evidence that school data use serves administrative reporting more than teaching and learning (Kostas et al., 2025) and that teacher data practice contracts toward assessment data despite broader definitions (Pastore, 2026). The consequence for instruction is tangible: the formative indications teachers most require, a pupil's trajectory of improvement or difficulty, remain dispersed across private records rather than informing timely adjustment, while standardized returns work against the individualization instruction demands. Interventions aimed only at individual teacher competence are therefore liable to fall short where the binding constraint is organizational; the leverage resides instead in aggregating data already held, so that attendance, read early, signals disengagement risk rather than an end–of–year verdict, and in rendering systems adaptable enough to restore agency to schools.

Two further implications concern theory and method. The accounts qualify the assumption, characteristic of proficiency–continuum models (Beck & Nunnaley, 2021), that data competence is mainly an individual attribute maturing with experience: difficulty was reported to follow the frequency of a procedure rather than the user's age or discipline, so infrequent tasks like end–of–term assessment unsettle even experienced staff, situating competence in routines and supports as much as persons. The contribution is thus to extend a multidimensional, role–differentiated account of educational data literacy toward a scale–differentiated one in which the organizational dimension carries the analytical weight, and to specify the mechanism, infrastructural mismatch, through which it conditions practice; the convergence of two institutional cases with the regional and national accounts lends the pattern a robustness a single site could not.

Conclusions

The SCHOOLWISE needs assessment advances three contributions. Conceptually, it reframes the underspecified role differentiation of educational data literacy as a scale–differentiated configuration and identifies an infrastructural mismatch, the distance between the data an actor regards as decision–relevant and the systems available to it, as a mechanism operating across scales and concentrated in the organizational dimension. Empirically, it demonstrates that this mismatch, together with a decoupling of responsibility from capacity, is discernible in how Italian school actors describe their own work, from the classroom to the national body. In practice, it indicates that data literacy in schools is advanced less by training individuals than by building organizational infrastructure and defining intermediate roles, a conclusion that will inform the project's subsequent instruments and training. These claims are preliminary and confined to a single national setting; their corroboration depends on the complete corpus and the survey data.

Keywords: Data literacy; Needs assessment; School practice; Role differentiation

Data Literacy and Data Use in Schools: Evidence from the DALIS Survey across Five European Education Systems

Sonia Marzadro – Davide Azzolini – Olga Gorodetskaya

Introduction

Schools increasingly operate in data–rich educational ecosystems, where digital platforms, assessments, administrative systems, and, more recently, AI–based tools generate unprecedented volumes of information. Yet, as highlighted in the literature on data–informed decision–making (Mandinach & Gummer 2016; Schildkamp & Lai 2013; Spillane 2012), the mere availability of data does not guarantee its meaningful use. Research consistently shows that educators' data literacy, organisational conditions, and professional cultures shape whether and how data support teaching, learning, and school improvement. This study builds on these frameworks to examine educators' competencies, attitudes, and practices related to data use across diverse European contexts, with particular attention to the emerging role of AI in schools.

Data Used

The analysis draws on the DALIS (Data Literacy in Schools) survey, developed within the Erasmus+ EVIDALI project. The survey was administered between 2025 and 2026 to a representative sample of primary, lower–secondary, and upper–secondary schools in five educational systems: Trentino (Italy), Liguria (Italy), Castilla y León (Spain), Madeira (Portugal), and Malta. Data were collected from teachers and school leaders through standardized questionnaires capturing:

- data literacy competencies;
- attitudes toward data;
- professional development experiences;
- school–level supports and collaborative practices;
- uses of AI–based tools in educational practice.

Methodology

The study adopts a comparative survey design based on stratified sampling across countries, school levels, and institutional characteristics. The DALIS instrument measures key dimensions of data literacy and data use, including frequency and types of data used, purposes of use in instructional and organisational decision–making, attitudes and self–reported competencies, participation in professional development, organisational supports, and patterns of AI adoption. The sampling strategy ensured representativeness across systems while adapting to structural differences in national education systems.

Quantitative analyses explore cross–country and within–school variation, as well as relationships among data literacy, professional development, school support, and actual data use.

Results and Discussion

Access and Use of Data

Results show a strong reliance on traditional data sources, particularly classroom assessments and behavioural records, while platform–based and administrative data are less frequently used. School leaders report higher levels of data use than teachers, especially regarding national assessment data. Barriers to data use are mainly related to perceived low usefulness of some data sources and lack of time.

Attitudes and Beliefs about Data

Attitudes toward data are generally positive: educators recognise its usefulness for teaching and learning despite acknowledging that data use is not always engaging. Concerns about privacy exist but do not translate into resistance. School leaders tend to show slightly more positive attitudes than teachers.

Data Literacy

Educators report high confidence in both their technical and pedagogical data–related competencies. Teachers' ability to read and interpret data is further supported by performance on the embedded quiz, indicating alignment between self–perceptions and demonstrated skills.

Professional Development

Only 4.5 out of 10 educators received data–related training during initial teacher education, while more than 6 out of 10 participated in such training during continuing professional development. A similar share—around 6 in 10—engage in self–directed learning on data use in their free time. Very few report having no training needs, though the needs expressed are highly diverse, spanning technical, pedagogical, and organisational dimensions.

School Data Use Strategies and Practices

Schools show generally good support levels for data use yet report limited training provision and relatively few structured projects dedicated to data–informed improvement. Teachers' perceptions of school support are broadly aligned with those of school leaders, but teachers are more uncertain about the specific measures implemented, suggesting communication or transparency gaps within schools.

Conclusions

The study offers one of the largest comparative studies of data literacy and data use among educators across multiple European regions. Results highlight the need to strengthen professional development, enhance school–level support structures, and develop sustainable data cultures that balance evidence–informed practices, professional judgement, and the responsible integration of AI technologies. Implications are relevant for teacher education, school leadership, and policy design, particularly in efforts to build schools' capacity to use data to improve teaching and learning.

Keywords: Data use; Data literacy; Evidence–informed decision–making

Combining Quantitative and Qualitative Evidence to Understand School Improvement in Vulnerable Contexts. Insights from the REVERS–ED Project

Lidia Bordanoba – Dragana Avramov – Maria Rita Mancaniello – Tiziana Chiappelli

Introduction

The growing availability of educational data has generated increasing interest in how schools and education systems use evidence to inform decision–making and improve educational outcomes. The use of data in education not only contributes to identifying problems and making informed decisions, but also to institutional learning through the validation or refutation of hypotheses, helping to avoid ineffective interventions and optimize resources in education policies (Schildkamp et al., 2016). Furthermore, at the school and classroom level, data from standardized tests, among other types of data, provide systematic information that makes it possible to identify learning difficulties from the earliest stages of education and to guide improvements in teaching and the development of essential basic competencies (OECD, 2023; Schildkamp, 2019). Without the data, teachers, families, and policymakers cannot compare which educational actions improve academic performance the most.

This paper presents findings from the Horizon Europe project *REVERS–ED. Trends on Educational Inequalities Over Time and Successful Interventions that Contribute to Reverse Them*. REVERS–ED focuses on analysing the long–term impacts of successful educational actions and interventions that are associated with improved outcomes for students experiencing disadvantage or vulnerability. To do so, the project integrates multiple sources of evidence, combining international statistics quantitative data, school–level longitudinal quantitative data, and qualitative evidence generated through the communicative methodology (Flecha & Soler, 2014). This mixed–methods approach enables a broader understanding of educational improvement by connecting measurable student

outcomes with the contextual, relational and organizational conditions in which they emerge. Through national and cross–school analyses, REVERS–ED provides insights that informs decision–making based on scientific evidence of social impact among researchers, educational practitioners and policymakers. Within this framework, the current paper examines how successful educational actions and interventions are associated with improvements in equity and inclusion across five European countries, highlighting the value of combining longitudinal and contextual evidence to understand and support school improvement in vulnerable contexts. The study is particularly relevant to current discussions on data use in schools because it examines how different sources of evidence can be integrated to understand educational improvement in contexts marked by poverty, migration, linguistic diversity, minoritized ethnicities and other forms of vulnerability.

Theoretical background

An extensive body of scientific literatura shows that Successful Educational Actions (SEA) implemented in over 15.000 diverse contexts worldwide are associated with improvements in both academic achievement and social cohesion when they are implemented with rigor. These actions are grounded in the theory of dialogic learning, which conceptualizes learning as a socially constructed process based on seven principles, egalitarian dialogue, cultural intelligence, transformation, instrumental dimension, creation of meaning, solidarity, and equality of differences (Flecha, 2000; Aubert et al., 2008). This framework emphasizes the central role of interactions and the inclusion of diverse voices as key drivers of both educational success and social transformation, moving beyond deficit–based perspectives (Flecha & Soler, 2013). Grounded in the principles of dialogic learning (Flecha, 2000), these actions are based on egalitarian dialogue and the active involvement of the educational community in overcoming inequalities. Indeed, SEA address multiple dimensions of school life that shape educational trajectories, including learning, participation, relationships, family and community engagement, wellbeing, and school climate.

SEA include Interactive Groups, Dialogic Literary Gatherings, inclusive family and community participation, and dialogic professional development among teachers (Valls & Kyriakides, 2013), among others.

Beyond the specific impacts of individual SEA, a growing body of research has identified a wide range of cumulative effects across different actions. These include improvements in instrumental learning as well as increased motivation and engagement, strengthened peer solidarity, reductions in conflict and violence, enhanced sense of belonging, and greater inclusion of traditionally marginalized groups (Flecha & Soler, 2013; López de Aguileta, 2019). Syntheses of evidence across SEA highlight how dialogic interactions simultaneously transform academic, social and emotional dimensions of learning, particularly in contexts of disadvantage (Valls & Kyriakides, 2013).

Among other SEA, Interactive Groups represent one of the most extensively researched SEA in the literature. Evaluative research across multiple contexts has shown Interactive Groups to be associated with improvements in literacy and mathematics attainment, reductions in disruptive behaviour and strengthened relationships between students from different backgrounds (Valls & Kyriakides, 2013, de Mello, 2012). Research on Dialogic Literary Gatherings (DLG), another SEA on which there is extensive research, has demonstrated their positive effects on reading comprehension, critical thinking, vocabulary development, motivation for reading and cultural self–confidence, with particularly pronounced benefits for students and families who have historically been positioned as outside the dominant literary culture (Pulido & Zepa, 2010).

Subsequent research has extended and critically examined the SEA framework across diverse European contexts. Studies conducted in Greece, Italy, Finland and other countries have explored how SEA are implemented in different institutional and cultural settings, identifying both the transferability of core principles and the importance of context–sensitive implementation (Soler–Gallart, 2015).

Importantly, the accumulated evidence also demonstrates the long–term sustainability of the dialogic approach. SEA have been maintained for more than 20 years across diverse educational systems,

showing consistent improvements in both academic and social outcomes over time (Flecha & Soler, 2013). This longitudinal perspective highlights that the dialogic approach is not only effective but also durable and scalable, including in schools serving highly vulnerable populations.

While the multiple social impacts of SEA are widely reported by the scientific literatura, REVERS–ED goes a step beyond and explores how their implementation is associated with changes in academic outcomes, participation, inclusion, and school climate in the long term in diverse vulnerable contexts across Europe. This approach highlights the importance of combining quantitative indicators with contextual and qualitative evidence in order to understand how schools create and sustain more equitable learning environments.

Methodology

The analysis draws on five national cross–case studies developed within REVERS–ED. The study includes 20 schools located in Greece, Ireland, Italy, Portugal and Spain. Schools were selected through a purposive maximum–variation strategy and shared two key characteristics: 1) they serve populations experiencing significant disadvantage and 2) they implement successful interventions, such as SEA, over an extended period.

The methodological design combines quantitative and qualitative evidence. Quantitative information includes longitudinal data of academic outcomes, attendance, progression and participation derived from external and/or internal evaluations, school monitoring systems and national assessment data where available. In some national contexts, this included longitudinal analysis of standardized assessment results and school–level monitoring data. These data were complemented by extensive qualitative evidence collected through the communicative methodology. The research involved interviews and focus groups with students, teachers, school leaders, families and other community members. This approach made it possible to analyze not only educational outcomes but also the interventions implemented and experienced within school communities.

Findings

Across the cases, quantitative indicators provide important information regarding attendance, progression and academic trajectories. The longitudinal analysis across the participating schools reveals a consistent pattern of improvement in students' academic outcomes, both in absolute terms and relative to their regional benchmarks. In all cases, schools initially performed at or below regional averages and progressively reduced these gaps or surpassed them over time. Notably, these improvements contrast with broader regional trends, which generally remained stable or showed slight declines during the same period.

From a data–informed school improvement perspective, the convergence of longitudinal school–level data and contextual evidence is particularly relevant. The replication of similar improvement trajectories across schools serving different vulnerable populations, educational levels, and territorial contexts suggests that these changes cannot be explained solely by local circumstances. Instead, the findings point to the sustained implementation of SEA and other successful interventions as a factor associated with positive educational trajectories in the long term. These results illustrate the value of longitudinal data not only for monitoring outcomes, but also for identifying practices that contribute to greater educational equity across diverse contexts.

The qualitative data allowed us to obtain a deeper understanding of the impacts of SEA. For instance, according to the perceptions of many of those interviewed, Interactive Groups improve coexistence and friendships. For example, in the case of a secondary education center, students placed strong emphasis on how Interactive Groups helped them make more friends and discover new people with whom they had not previously interacted. Regarding Dialogic Literary Gatherings, many students stated that for the first time they have been able to get to know some of their classmates in depth, as previously they may not have heard them speak or had not heard them talk about topics that are important to them.

The inclusion of qualitative evidence made visible dimensions such as emotional safety, recognition, agency, participation and belonging, which repeatedly emerged as central mechanisms associated with positive educational trajectories.

Discussion

The findings contribute to current debates on data use in schools by suggesting that evidence–informed improvement in vulnerable contexts requires broader conceptions of educational data. The cross–case analysis identified a set of recurring conditions associated with the successful implementation of SEA in highly vulnerable contexts. Longitudinal evidence across the analyzed schools points to the importance of rigorous implementation grounded in dialogic learning principles, high expectations for all students, active family and community participation, extended learning opportunities, and supportive environments based on solidarity and inclusion. The findings also highlight the role of sustained implementation over time, allowing schools to monitor educational trajectories and consolidate improvements.

Importantly, across diverse educational levels, territories, and student populations, similar patterns of improvement were observed, suggesting that the social impacts of SEA are transferable across contexts.

In this regard, the findings reinforce that the dialogic approach generates not only immediate improvements but also sustained transformations over time. Schools that have implemented SEA longitudinally show stable and cumulative gains, indicating that these practices continue to produce impact across cohorts and changing conditions over extended periods (Flecha & Soler, 2013).

The results indicate that educational inequalities are effectively addressed through interventions supported by scientific evidence of social impact and embedded within educational communities that promote dialogue, collaboration, and shared responsibility for student success. Rather than relying on compensatory approaches, these schools combined longitudinal evidence with inclusive practices to expand participation, reduce low achievement, and improve educational outcomes, demonstrating the impact of SEA as scalable for equity–oriented school improvement.

Furthermore, sustainability emerges as closely linked to the institutionalization of dialogic practices within school communities. When SEA become embedded in everyday interactions, organizational routines, and community engagement, they enable the long–term maintenance and consolidation of improvements, particularly in vulnerable contexts (Flecha & Soler, 2013).

Conclusions

The REVERS–ED project demonstrates the value of combining multiple evidence sources to analyze educational improvement in contexts of disadvantage. The findings indicate that successful educational actions and interventions that reverse educational inequalities are associated not only with improvements in academic performance but also with transformations in promoting dialogic relationships among students and the wider community, a decrease in violence and conflicts, and overall increased social cohesion and inclusion.

For educational policy and practice, these results point to the need for evidence systems that integrate quantitative longitudinal and qualitative information, enabling a more comprehensive understanding of educational trajectories and the impact of school interventions over time. Such an approach contributes to more meaningful forms of decision–making based on scientific evidence of social impact capable of supporting equity–oriented school improvement.

Keywords: Longitudinal data; Reversing inequalities; Successful educational actions; Mixed methods

The Perception of Standardized Testing Between Teaching, Accountability, and Emotion: A Longitudinal and Multi–Perspective Analysis at the "Istituto Comprensivo" of Praia

Emilia Ciccia – Marilena Salsano

Introduction

National standardized tests provided by the National Institute for the Evaluation of the Education and Training System (INVALSI) constitute a fundamental pillar of accountability and monitoring policies within the Italian school system. Nonetheless, the introduction of these assessments continues to generate intense cultural and pedagogical debates among stakeholders in the school system. Scientific literature highlights that standardized assessment is not a neutral process. National and international studies detect the presence of the *test anxiety* phenomenon, which can undermine not only students' psychological well–being but also the very reliability of the results (cf. Poliandri et al., 2019; Quintano et al., 2021).

Furthermore, research often highlights an epistemological discrepancy between the institutional purpose of the tests (system evaluation and support for instructional design) and the perception of the school community, which sometimes interprets the tests as tools for bureaucratic control or meritocratic classification of students (Corsini, 2018; Grimaldi & Serpieri, 2013; Roccia, 2021).

This study fits into the line of research aimed at exploring the evaluative climate. The primary objective is to analyze the evolution of the perception of INVALSI tests within the *Istituto Comprensivo* of Praia, by directly comparing the viewpoints of three key actors: students, teachers, and parents. The investigation aims to verify whether the educational–informative actions undertaken by the school have mitigated performance anxiety and enhanced the formative use of empirical data.

Data Used

The research is based on empirical data extracted from six distinct datasets distributed over two school years (2024–25 and 2025–26), relating to the *Istituto Comprensivo* of Praia. The samples include fifth–grade students from Primary School and eighth–grade (*terza media*) students from Lower Secondary School, together with their teachers and parents. The research instruments used for multi–perspective triangulation consist of three structured questionnaires created *ad hoc*, aimed at capturing the perception, attitude, and emotional experience of the three main actors in the school community.

The questionnaires were administered via digital forms and share an architecture predominantly based on agreement scales (Likert) and multiple–choice questions, organized around specific macro–dimensions:

- **INVALSI Student Questionnaire:** This instrument explores the direct experience of the pupils (fifth grade of primary school and third grade of secondary school), focusing on preparation and the emotional impact of the tests.
  - *Educational/Preparatory Dimension:* Investigates whether exam simulations were carried out during the year and through which methods/educational supports provided by the teachers.
  - *Emotional Dimension:* Precisely measures the state of mind with which students faced the tests, asking them to express their level of agreement regarding contrasting feelings such as: Anxiety and Worry; Curiosity and Satisfaction/Challenge; Calmness or Indifference.
- **INVALSI Parent Questionnaire:** This questionnaire is aimed at measuring the level of awareness among families, the support perceived from the school, and the psychological reflection on their children.
  - *Level of Information:* Assesses the degree of knowledge about the tests and whether the family received adequate clarifications from teachers regarding the purpose of the assessments.

- *Perception of Educational Support:* Monitors whether, from the parent's point of view, their child was well–supported at school in preparing for the tests.
- *Well–being and Pressure:* Detects the domestic psychological climate, asking whether the child experienced the tests calmly or if, conversely, the testing framework generates too much pressure on students.
- *Institutional Value:* Measures agreement on the usefulness of standardized tests for understanding children's actual level of preparation.

- **INVALSI Teacher Questionnaire:** This is the most complex tool from a macro–evaluative and logistical standpoint, structured to understand the cultural positioning of the teaching staff regarding accountability.
  - *Valuational and System Dimension:* Explores teachers' agreement or disagreement with polarized statements, defining the tests either as a useful tool for evaluating the school system / designing specific training courses for teachers, or as data useful for daily educational improvement. It also captures if they are perceived negatively (e.g., "a waste of time", "used to judge students", or "perceived as a control over teachers' work").
  - *Logistical and Climate Dimension (Emerging from the overall structure):* Measures the integration of institutional Reference Frameworks and the effectiveness of organizational–educational supports prepared by the school (including familiarization with Computer–Based and paper–based tests). This dimension is directly linked to school governance strategies, with particular reference to the impact of sharing and guided reading of school results coordinated by the INVALSI Coordinator (*Referente*).

Methodology

A quantitative, multi–perspective (stakeholder triangulation), and longitudinal research design was adopted. The methodology is structured along three fundamental analytical directions:

1. **Stakeholder Triangulation:** Responses provided simultaneously by students, teachers, and parents on overlapping macro–factors (emotional dimension, perceived utility, and test preparation) were compared.
2. **Longitudinal Analysis:** The shift in macro–trends between the 2024–25 school year and the 2025–26 school year was evaluated to measure the impact of school policies.
3. **Segmentation by Education Level:** Responses from fifth–grade primary school classes were isolated from those of third–grade secondary school classes to investigate the transition effect and the anxiety load associated with the proximity of the State Examination.

Data analysis relies on descriptive and inferential statistics to highlight significant convergences and divergences between the groups.

4. Results and Discussion

The sample data reveal a clear polarization and major inter–group discrepancies:

- **Emotional Dimension and Psychological Climate:** Comparative analysis reveals a significant disconnect between the actual emotional experience of students and the perception of adults (teachers and parents). Contrary to initial expectations linked to the proximity of the State Examination, fifth–grade Primary students manifest the most marked emotional distress and peaks of anxiety and worry. In the Student Questionnaire dataset, 50% of fifth–grade pupils report medium–high levels of anxiety and a staggering 56.2% say they are worried, compared to significantly lower percentages in the third year of secondary school (39.4% anxiety and 36.8% worry, respectively). Eighth graders show a distinctly higher capacity for rationalization and "detachment" (37–38% declare themselves not anxious at all, nearly double the primary school rate). Conversely, teachers tend to describe their own emotional state as calm and enthusiastic, struggling to fully intercept this strong emotional vulnerability in younger primary school children, while parents position themselves in an intermediate state of high apprehension.

- **Teaching Dimension and Preparation:** Longitudinal analysis (a.y. 2024–25 vs. a.y. 2025–26) demonstrates that the systematic administration of exam simulations and information transparency regarding test goals act as an effective mitigation factor for student anxiety. In fifth–grade Primary, a strong shift is recorded from acute anxiety to moderate or absent anxiety, indicating good guided familiarization with both paper and digital tools. In third–grade Secondary, anxiety levels remain overall low and stable. Teachers report a robust use of INVALSI Reference Frameworks within curricular planning. Excellent consistency is found between the activities declared by teachers and those perceived by students in the second year of data collection, confirming a progressive familiarization with the Computer–Based Testing (CBT) mode and a progressive maturation and institutionalization of *data literacy* among the teaching staff.
- **System Evaluation vs. Individual Judgment:** A durable cultural resistance persists among families. Parents frequently express reservations about the ability of standardized tests to measure students' actual skills and human value. On the contrary, teachers demonstrate an ability to separate the utility of INVALSI data as a system and school self–evaluation indicator from the fear of an inspectional or punitive control over their individual work.

Conclusions

The study demonstrates that educational familiarization and transparency significantly reduce test anxiety, normalizing the perception of standardized testing. To break down the punitive or merely fiscal conception of assessment that still persists within the family nucleus, it appears essential for schools to implement synergistic and transparent communication strategies toward parents. The longitudinal data collected confirm that the formative and non–bureaucratic use of INVALSI data by teachers consolidates equity practices and fosters continuous improvement in the educational offerings of the *Istituto Comprensivo* of Praia.

Keywords: INVALSI tests; Standardized assessment; Data triangulation; Longitudinal analysis; Emotional well–being

Cross–National Patterns of Teachers' Professional Development and Classroom Assessment Practices: Evidence from TALIS

Francesco Ottaviani – Serafina Pastore

Introduction

Educational reforms have increasingly emphasised quality, measurable outcomes, accountability and the comparison of education systems. Within this context, assessment has acquired a dual role: it is both a pedagogical process through which teachers support student learning and a policy instrument used to monitor and improve educational quality.

In this study, the concept of teacher assessment literacy provides the theoretical framework for understanding the knowledge, skills, beliefs and practices involved in teachers' assessment work. While early definitions focused mainly on the technical ability to design, administer and interpret assessments, recent approaches describe assessment literacy as a multidimensional and context–dependent construct shaped by educational policies, professional cultures, school organisation and teachers' beliefs (Coombs & DeLuca, 2022; DeLuca et al., 2020; Pastore & Andrade; Willis et al., 2013; Xu & Brown, 2016).

Continuous professional development (CPD) may support the development of assessment literacy by helping teachers revise established classroom routines and adopt formative and student–centred practices. However, evidence on its effectiveness remains mixed, and its contribution may depend on

programme content, duration, coherence, opportunities for collaboration and connections with classroom practice (Looney et al., 2018; Sims et al., 2025). Moreover, most existing studies focus on individual programmes or national contexts, providing limited evidence on changes across education systems and over time.

This study examines assessment–focused professional development and teachers' classroom practices across European education systems. It addresses two main research questions:

RQ1. What trends can be identified in teachers' participation in assessment–focused professional development, their reported learning needs and their classroom assessment practices?

RQ2. Do teachers who participate in assessment–focused professional development report different assessment practices from those who do not?

Data used

The study uses teacher–level data from the OECD Teaching and Learning International Survey — TALIS 2013, 2018 and 2024. The analysis focuses on lower secondary education, corresponding to ISCED level 2, and includes 25 European education systems participating in at least two of the three survey waves.

TALIS adopts a stratified two–stage probability sampling design, selecting schools and then teachers within participating schools. The analysis applies the official teacher and replicate weights to account for unequal selection probabilities, non–response and the complex survey design.

Six harmonised indicators cover three dimensions of teachers' assessment–related professional activity:

1. participation during the previous 12 months in professional development covering student assessment;
2. a high reported need for further professional learning in student assessment;
3. a high reported ability to use a variety of assessment strategies;
4. always providing oral or written feedback indicating areas for improvement;
5. always asking students to assess their own progress;
6. always observing students while they work and providing immediate feedback.

The indicators have been harmonised across survey waves according to item wording, response scales and conceptual comparability. They are expressed as percentages of teachers selecting the relevant response category.

Methodology

The analysis proceeds in two stages.

First, it examines system–level trends in the six indicators. For each education system and survey wave, the study estimates weighted percentages and compares subsequent observations with the earliest available year. When 2013 data are available, 2013 is used as the baseline; otherwise, 2018 serves as the reference year.

Second, the study conducts within–system comparisons between teachers who participated in assessment–focused professional development during the previous 12 months and those who did not. Separate linear regression models are estimated for each education system, survey wave and classroom–practice indicator. Group differences are reported in percentage points.

All estimates use the official TALIS weights and balanced repeated replication weights. Standard errors are calculated through the OECD REPEST procedure in Stata (Avvisati & Keslair, 2024; OECD, 2019).

The within–system estimates should not be interpreted as causal effects. Participation in CPD is not randomly assigned and may depend on teachers' motivation, previous competencies, school characteristics and access to training. The analysis therefore identifies associations and patterns of differentiation rather than the effectiveness of a specific intervention.

Results and discussion

The results show substantial differences across European education systems and heterogeneous trajectories between 2013 and 2024.

Participation in assessment–focused professional development initially ranges from below 30% to above 80%. Systems starting from comparatively low levels generally record the largest increases, whereas systems with higher initial participation show more stable, moderate or declining trajectories. These developments indicate partial rebalancing, but not complete convergence: access to assessment–focused professional learning remains uneven across Europe.

The need for further professional development reported by teachers follows differentiated trajectories. In some systems, reported needs increase, while in others they remain stable or decline. These needs should not necessarily be interpreted as evidence of skill deficits. They may reflect greater awareness of the complexity of assessment and exposure to new professional expectations.

Teachers' reported ability to use a variety of assessment strategies displays also a similar pattern. Improvements are more frequent among systems with lower baseline levels, whereas several systems with initially high levels remain stable or record declines. The overall trend therefore suggests rebalancing rather than uniform improvement.

The classroom–practice indicators evolve differently. Providing oral or written feedback to identify areas for improvement shows the clearest expansion, particularly where it was initially uncommon. In contrast, student self–assessment is adopted more slowly and remains relatively uncommon in several systems. One likely explanation for this is that its implementation requires teachers to clarify learning goals, communicate assessment criteria, and provide students with structured opportunities to reflect on their learning (Panadero et al., 2016).

Other classroom assessment practices, such as observing students while they work and providing immediate feedback, have changed more modestly. As this approach was already widely adopted in earlier stages, it appears to be approaching a state of stability or reconfiguration rather than further diffusion.

Italy, for example, illustrates the multidimensional character of these changes. Participation in assessment–focused CPD increases substantially, while the proportion of teachers reporting a high need for additional training declines. This national school system also records marked increases in systematic feedback and student self–assessment. However, the proportion of teachers reporting a high ability to use a variety of assessment strategies decreases moderately, while immediate feedback during classroom tasks increases only slightly. Participation, perceived needs, self–reported capabilities and specific classroom practices therefore do not necessarily move in the same direction.

The within–system comparisons reveal a consistent positive association between assessment–focused professional development and teachers' reported practices. Across the three waves, participating teachers generally report greater assessment capability and more frequent use of formative practices than non–participants. In 2013, these differences are usually small and statistically significant in a limited number of education systems. They become more widespread in 2018 and are both larger and more pervasive in 2024.

In the most recent wave, participation in assessment–focused CPD is associated with differences of approximately:

- 6–12 percentage points in teachers' reported ability to use a variety of assessment strategies;
- 3–13 percentage points in the systematic provision of oral or written feedback;
- 3–10 percentage points in the use of student self–assessment;
- 3–10 percentage points in the provision of immediate feedback while students are working.

The largest gaps often concern practices requiring more complex forms of assessment literacy, particularly systematic feedback and student self–assessment. These practices require teachers not only to apply specific techniques, but also to interpret evidence, communicate objectives and actively involve students in assessment.

The results suggest that professional development is becoming an important axis of differentiation among teachers. Participation is associated with greater use of formative practices, but access to relevant professional learning may remain unequal. CPD may therefore support improvements

among participating teachers while also widening differences between those who have and have not accessed such opportunities.

Selection mechanisms remain an important alternative explanation. Teachers who are more motivated or already more assessment–literate may be more likely to participate. Schools with stronger professional learning cultures may simultaneously provide greater access to CPD and more favourable conditions for formative assessment. Nevertheless, the increasingly widespread gaps indicate a close relationship between professional learning and the enactment of assessment in the classroom.

Conclusions

The development of teacher assessment literacy follows a differentiated rather than fully convergent trajectory across European education systems. School systems starting from lower levels of CPD participation or lower prevalence of practices often record the greatest progress, but substantial heterogeneity remains in 2024.

Different assessment practices also evolve at different rates. Systematic feedback becomes more widespread, whereas student self–assessment remains less consolidated. Immediate feedback during classroom tasks changes only modestly, partly because it was already relatively common.

Within education systems, teachers participating in assessment–focused CPD consistently report greater use of assessment strategies and formative practices. These differences are particularly widespread in 2024, suggesting that professional development is increasingly associated with differentiation in classroom practice.

Policies should therefore focus not only on increasing overall participation in CPD but also on ensuring equitable access to sustained, coherent and classroom–embedded professional learning. Developing assessment literacy requires more than introducing new tools: it involves teachers' beliefs, school organisation, professional collaboration and the purposes for which assessment evidence is used.

Future research should examine the content, duration and quality of professional development and the role of school leadership, professional standards and teacher education policies. It should also investigate whether differences in teacher assessment literacy and classroom practices are associated with differences in student learning outcomes.

Keywords: Continuous professional development; Classroom assessment; Assessment data literacy; TALIS

SESSION 8. EVALUATING UNIVERSITY PATHWAYS IN ITALY: INTEGRATED DATA, EDUCATIONAL CHOICES, STUDENT SUPPORT AND LABOUR MARKET TRANSITIONS

ORGANIZERS: UNIVERSITY OF PIEMONTE ORIENTALE – UNIVERSITY OF BOLOGNA
COORDINATORS: CARMEN AINA – GIUSEPPE PIGNATARO
5TH NOVEMBER: 8.30 A.M. – 10.30 A.M. {AULA A1.1 – RESEARCH 8}

Linking administrative data on university careers and labour market outcomes to study regional mobility in Italy

Dalit Contini – Sonia Della Monaca – Noemi Ferrara – Francesca Prever – Andrea Priulla

Motivation and research question

Regional mobility among university students and graduates has become a central issue in the Italian debate on territorial inequalities. The movement of young people from Southern regions and the Islands towards metropolitan and economically stronger areas of the Centre–North, and in some cases abroad, is often discussed as a form of brain drain, with important consequences for regional development, the distribution of human capital, and education and labour–market policies. Understanding when mobility occurs, whether it is temporary or persistent, and whether it leads to qualified employment is therefore crucial for a better understanding of territorial inequalities.

Despite the relevance of the issue, empirical evidence on interregional mobility after university remains partial. The main source currently available in Italy is AlmaLaurea, which combines university administrative records with surveys administered to graduates of member universities at graduation and one, three and five years after graduation. AlmaLaurea provides rich and widely used information on degree programmes, study experiences, and early labour–market outcomes, and it remains an essential benchmark for research and policy monitoring. However, it has a few limitations. First, information is collected at discrete points in time and does not allow the continuous reconstruction of individual sequences of employment, unemployment, inactivity and regional moves. Second, follow–up surveys are affected by attrition, which increases over time after graduation and is not random; in particular, the work episodes of bachelor graduates are not recorded when they enrol in a master's programme. Third, AlmaLaurea observes graduates only, whereas students who are still enrolled or who leave university without completing a degree are not included in the survey.

Italy also has two families of administrative archives that could substantially improve the study of transitions from higher education to work. The National Student Register (Anagrafe Nazionale Studenti, ANS) follows the university careers of students enrolled in Italian universities, from prior schooling to enrolment, transfers, credits, graduation or drop–out. WHIP (Work and Health Italian Panel), developed from INPS data, reconstructs labour–market histories for a representative large sample of workers in Italy, including timing and location of employment episodes and detailed occupational information. If linked, these sources would allow researchers to follow individuals from region of birth and upper–secondary school, through university enrolment, to the region and type of work observed after university.

This long abstract presents a privacy–preserving experiment linking ANS and WHIP. The contribution is methodological and substantive. Methodologically, we show how administrative archives produced by different institutions can be integrated without direct personal identifiers through a double–blind linkage design and biographical consistency checks. Substantively, we use the linked one–to–one dataset to provide preliminary evidence on internal mobility between

university and work. The focus is on how mobility differs by degree level, drop–out status, area of study and occupational qualification, with particular attention to the unequal capacity of Italian macro–regions to retain, attract or regain young human capital.

Data infrastructure and linkage strategy

This study builds on the experimental integration of two administrative data infrastructures: MOBYSU.IT, derived from the National Student Register, and WHIP.

MOBYSU.IT is a research database based on the *Anagrafe Nazionale Studenti* (ANS), the national administrative register that records the careers of students enrolled in Italian universities. It was created within a research protocol involving the Ministry of University and Research, INVALSI and several Italian universities, with the aim of studying university trajectories and student mobility on a national scale. Unlike single–university archives, MOBYSU.IT makes it possible to follow students across institutions and regions, observing enrolment, transfers, changes of degree programme, credits, graduation and drop–out. It also includes information on demographic characteristics, upper–secondary education, region of school attendance, type and grade of diploma, university attended, field of study and timing of academic progression.

For this experiment, we use cohorts of first–time entrants to Italian universities between the academic years 2009/10 and 2016/17, observed until June 2019. The sample is restricted to students who entered university by age 22, excluding those enrolled in online universities and those in single–cycle degree programmes. This selection focuses on standard university–to–work transitions for bachelor's and master's students and avoids programmes leading to regulated professions that are only partially observed in WHIP. The resulting ANS/MOBYSU.IT sample includes 96,968 individuals, selected according to the same birth–date sampling scheme used in WHIP.

WHIP is a longitudinal administrative database covering a representative sample of workers in Italy. It is built from several administrative sources, including INPS records and the Ministry of Labour's Compulsory Communications, which report the activation, transformation and termination of employment contracts. For this study, WHIP provides information on employment spells, region of work, contract timing and occupational classification. We use individuals from the same birth cohorts selected in MOBYSU.IT, born in Italy and with at least one ISEE declaration, since the municipality of birth needed for linkage is available in WHIP only through this information flow.

The linkage between MOBYSU.IT/ANS and WHIP is performed without direct personal identifiers, through a privacy–preserving, double–blind design based on sex, full date of birth and municipality of birth. One team identifies records sharing the same linkage key, while a separate team analyses the resulting educational and employment biographies using only aggregated birth information, so that no team has access to both disaggregated identifying variables and substantive trajectories.

The linkage procedure produces different configurations: one–to–one, one–to–many, many–to–one, many–to–many and no match. Exact agreement on sex, date and municipality of birth does not guarantee that two records refer to the same person, especially in larger municipalities. For this reason, initial matches are refined through biographical consistency checks based on the temporal sequence of education and work. Before university, the region of upper–secondary school is compared with the worker's region of residence or early employment. During university, the region of the university is compared with the region of residence or employment. After university, occupational information is checked against degree attainment and field of study, excluding implausible combinations such as regulated or highly specialised professions without a compatible degree.

After these procedures, the initial sample is divided into 16,777 one–to–one matches, 10,117 one–to–many matches, 16,252 many–to–one matches, 15,453 many–to–many matches and 38,369 no matches. The present analysis focuses on the one–to–one linked dataset, which provides the most reliable basis for preliminary evidence on mobility between university and work. Non–unique matches are left for future developments based on probabilistic assignment and multiple–imputation strategies.

The one–to–one dataset contains 16,777 students with linked labour–market histories. It is not fully representative of all students with work experience: the design favours individuals born in smaller municipalities, where the linkage key is more discriminating, and workers visible in regular employment records and in ISEE–related information. However, exploratory comparisons between one–to–one matches and non–unique matches show limited differences in most educational characteristics. Apart from municipality size, the probability of belonging to the one–to–one subset is only weakly associated with gender, field of study or university pathway. Males are about three percentage points more likely to be in the one–to–one group; some disciplinary differences are statistically significant but substantively small. These results suggest that the one–to–one sample can be used for exploratory mobility analyses, provided that selection mechanisms are explicitly acknowledged and sensitivity analyses are planned.

Preliminary evidence on mobility from university to work

The linked archive makes it possible to observe a sequence that is usually unavailable in a single source: region of birth, region of upper–secondary school, region of university, degree completion or drop–out, and the regions and occupations of subsequent employment episodes. This structure is particularly useful for studying internal mobility because it distinguishes mobility for study from mobility for work and because it includes both graduates and non–completers. In the analyses below, mobility is defined at the level of macro–regions (North–West, North–East, Centre, South and Islands). For graduates, employment episodes are considered from the year of degree completion onwards; for students who drop out, the reference point is the last university attended. Occupations are grouped by the first digit of the ISCO classification into high–qualified (codes 1–3), medium–qualified (4–6) and non–qualified jobs (7–8).

The first descriptive result is that degree level matters. Among master graduates, 30% experience at least one post–graduation move away from the macro–region of the university, while 70% remain in the same macro–region. Among bachelor graduates, mobility is lower: 23% experience at least one move and 77% remain. This difference suggests that higher educational attainment broadens the relevant geographical labour market.

Mobility is also selective with respect to job quality. Employment transitions across macro–regions are more frequent in medium– and high–qualified occupations than in non–qualified ones. This pattern is consistent with the interpretation of mobility as part of a strategy of human–capital investment: moving is more likely when it offers access to more dynamic labour markets, better matching between education and occupation, or professional opportunities that are not available in the region of study. At the same time, mobility is highly asymmetrical across territories. The Centre and the North retain most of their graduates, with stay rates generally above 70%, especially for high–qualified positions. By contrast, the South and the Islands lose a much larger share of their graduates to other macro–regions.

The clearest imbalance concerns high–qualified jobs among master graduates from Southern universities. More than one contract out of two in this segment – 53.9% – is located outside the macro–region of study, with a marked orientation towards the North–West. The North–West emerges as the strongest pole of attraction not only for high–qualified jobs, but also for some lower–qualified employment episodes. For example, among master graduates from the Islands who hold non–qualified jobs, a large share of such contracts is recorded in the North–West. This evidence indicates that South–to–North mobility is not simply a flow of highly skilled graduates into high–quality jobs; it may also include transitions in which individuals leave weaker labour markets even without obtaining highly qualified positions.

Return mobility adds an important nuance. Among bachelor graduates who moved for study, 41.4% return to their macro–region of origin at least once during the observed labour–market career. Return rates are much lower for students from the South and the Islands than for those from the Centre and the North. Among movers from the South and the Islands, return rates are respectively 43.4% and 47.3% for bachelor graduates, but fall to 26.0% and 29.3% for master graduates. In the

Centre and the North, by contrast, return rates are roughly between 60% and 80% for bachelor graduates and between 43% and 72% for master graduates. These figures suggest that mobility out of Southern areas is less likely to be temporary, especially after completion of a master degree. This asymmetry is crucial for policy: regions that already face weaker labour–market conditions not only lose students for university but also have a lower probability of reabsorbing them after graduation.

Students who leave university without a degree follow a different pattern. Drop–outs tend to remain in the macro–region of the university much more often than graduates do: in most areas, stay rates exceed 80–85%. Their limited mobility is also less clearly associated with access to high–qualified work. In the South, the capacity to retain or attract qualified employment for drop–outs is particularly weak; among students who abandoned a bachelor programme after studying in the South, less than 5% of contracts are in high–qualified occupations, while in Northern areas the corresponding share exceeds 10%. The few observed moves among drop–outs are again oriented mainly towards Northern macro–regions, but the phenomenon is smaller and less structured than among graduates.

The comparison between graduates and drop–outs therefore shows that university completion changes the meaning of mobility. For graduates, especially master graduates, mobility appears selective and more closely linked to medium– and high–qualified labour–market segments. For non–completers, mobility is more localised, less frequent and less connected with high–quality occupational insertion. At the same time, drop–out can be associated with return: the return rate to the macro–region of origin is 47.4% among students who abandoned a master programme and 43.8% among those who abandoned a bachelor programme. This suggests that some failed or interrupted university trajectories are temporary mobility experiences that end with a move back to the area of origin.

Discussion and policy implications

The experiment shows that administrative data linkage can contribute to the evaluation of education and labour policies in at least three ways. First, it allows researchers to follow individual trajectories beyond graduation and beyond the survey interview calendar. This is important for evaluating whether university mobility produces durable territorial reallocation, temporary moves or return paths. Second, it makes visible a population usually excluded from graduate surveys: students who leave university without completing a degree. Their trajectories are essential for assessing the broader social consequences of university access, mobility costs and local labour–market opportunities. Third, the availability of detailed occupational information makes it possible to distinguish mobility that leads to qualified employment from mobility that reflects weaker local labour demand without clear occupational upgrading.

The findings are preliminary and must be interpreted in light of the linkage design. The one–to–one subset is affected by demographic geography, since the linkage key is more precise for people born in smaller municipalities, and by administrative coverage, since WHIP observes regular employment relationships and the linkage requires ISEE–related information on the WHIP side. No–match cases cannot be interpreted mechanically as absence of work: individuals may work abroad, work in sectors not covered by WHIP, perform occasional activities, or lack the information required for linkage. Future work should therefore validate the one–to–one matches against local deterministic linkages where available and develop probabilistic strategies for using non–unique matches while explicitly modelling uncertainty.

Despite these limitations, the linked data already reveal patterns that are directly relevant for the Italian policy debate. Mobility is not simply an individual choice or a neutral redistribution of human capital. It is structured by degree level, occupational opportunities and regional labour–market capacity. The South and the Islands experience the most problematic configuration: high rates of outward mobility for study and work, lower rates of return among graduates, and weaker access to qualified employment for those who abandon university. Policies aimed at reducing territorial inequalities should therefore address both sides of the transition: the conditions that push students

to move for university and the labour–market conditions that determine whether they can return with appropriate occupational opportunities.
More generally, the value of this experiment lies not only in the specific estimates reported here but in the research infrastructure it makes possible. A stable privacy–preserving linkage between education and employment archives would allow continuous monitoring of university–to–work transitions, regional retention, return mobility, occupational matching and the consequences of drop–out.

Keywords: Graduate migration; Administrative data; Record linkage; University–to–work transition

STEM Education–to–Work Transition: A LinkedIn Correspondence Audit on Gender Disparities

Elena Grimaccia – Martina D'Amato – Simone Dei Pieri– Iacopo Moreschini – Fabio Ferri Daniele Montepara

Introduction

The digital transformation of labour markets has opened new possibilities for socio–economic research, augmenting traditional data sources with large–scale, real–time information drawn from professional networking platforms (Einav and Levin, 2014; Varian, 2014). This study exploits that transformation to examine gender disparities in access to Science, Technology, Engineering and Mathematics (STEM) careers in Italy. Although women now represent 59.7% of Italian university graduates and outperform men academically (AlmaLaurea, 2023), they account for only 39% of STEM graduates and 34.1% of scientists and engineers, below the EU average of 41% (Eurostat, 2025). This under–representation reflects persistent occupational segregation: reducing it bears directly on two of the United Nations Sustainable Development Goals on quality education (SDG 4) and gender equality (SDG 5), whose targets explicitly link equitable access to education with women's full participation in the labour market (United Nations, 2015).
Women's under–representation in STEM matters because it shapes the returns to human capital, the equality of educational opportunity and the transition from education to work (Pignataro, 2012). While supply–side mechanisms (educational choices, career preferences) have received considerable attention, demand–side mechanisms, and recruiter behaviour in particular, remain under–examined because of data and methodological constraints. The prior literature shows that the direction of hiring bias is strongly occupation–specific: in male–dominated fields, including many STEM occupations, recruiters may actually favour female candidates, possibly reflecting the growing reach of corporate diversity initiatives (Galos and Coppock, 2023). Evidence on gender–based recruiter responsiveness in the Italian digital labour market is scarce and largely observational rather than experimental (Berte et al., 2023; Boscaino et al., 2026).
We address this gap by adapting the classical correspondence audit design to a digital environment. To our knowledge, this is the first application of the audit experimental methodology to the Italian digital labour market for STEM occupations, providing causal rather than correlational evidence. The study answers two research questions: whether perceived gender affects (i) the probability that an application is viewed by a recruiter, and (ii) the probability of receiving a positive response. Establishing whether qualified women face a demand–side penalty at the point of entry is a precondition for well–targeted equity policy, since it determines whether corrective effort is best directed at recruitment or, instead, further upstream in the STEM education pipeline.

Data

We collected an original dataset on LinkedIn between 15 and 30 January 2025. Using an Apify scraper with four search queries (Junior Data Analyst, Data Scientist, Junior Backend Developer, Software

Developer), we retrieved 939 postings and filtered them for relevance, obtaining 197 job postings to which both fictitious candidates applied simultaneously, for a total of 394 paired applications. After retaining only postings where both applications could be submitted, 167 valid paired observations entered the analysis. Postings concentrated in Northern Italy (56.3%), led by Lombardy (17.8%); Data Scientist roles were the most frequent category (34.0%), and 90.4% of postings offered full–time entry–level employment, matching the profile of recent graduates.

Two fictitious profiles were created: Chiara Romano (female) and Francesco Russo (male). Their CVs were identical in structure, length and content, both reporting a Bachelor's degree in Computer Science (L–31) from the University of Milano–Bicocca, equivalent technical skills (PHP, SQL, Python, Power BI, GIT) and soft skills, the same prior work and volunteer experience, and identical language proficiency. The profiles differed only in gender–signalling attributes: name, AI–generated profile photo and cosmetic formatting (font and colour). The recorded outcomes were whether the application was viewed, whether any response was received, and whether the response was positive (e.g. an interview invitation). As a transparency note, we observed that AI image generation reproduced gendered aesthetic stereotypes, rendering female faces as more decorative and less professional under identical settings, an illustration of algorithmic visual bias relevant to digital recruitment (Nicoletti and Bass, 2023).

Methodology

The study follows the classical correspondence audit methodology (Riach and Rich, 1987; Bertrand and Mullainathan, 2004), adapted to a platform–mediated setting (Lahey and Beasley, 2009). It implements a fully randomised paired experiment (Cox, 1958) in which only the perceived gender of the applicant varies while all other characteristics are held constant, supporting clean causal inference on recruiter behaviour. Because both profiles were submitted to each posting under structurally identical conditions, any residual algorithmic differentiation operates exclusively through gender signals and therefore forms part of the gender effect being measured rather than a confound.

We estimate four logistic regression models (binomial GLM with logit link; Hosmer and Lemeshow, 2000), modelling visualisation (Models 1 and 3) and positive response (Models 2 and 4) as a function of a Female indicator, with and without geographic–area controls (North, Centre as reference, South and Islands). Because observations are paired, we additionally exploit McNemar's test for matched binary outcomes (McNemar, 1947). Given the small number of discordant pairs ($b + c < 20$), we rely on the exact binomial version of McNemar's test, which computes exact probabilities under the null hypothesis without large–sample approximation and is therefore valid at these sample sizes (Agresti, 2013). Robustness is assessed in two ways: a subgroup analysis separating analytical roles (Junior Data Analyst and Data Scientist; $n = 94$) from technical roles (Junior Backend Developer and Software Developer; $n = 73$), and a non–parametric bootstrap (10,000 replications) that resamples the discordant pairs with replacement to gauge the stability of the exact–test results (Efron and Tibshirani, 1993).

Results and discussion

Female applications were viewed more often than male ones (43.7% vs 35.9%) and received markedly more positive responses (6.5% vs 1.8%). In the matched–pairs distribution for visualisation, 17 discordant pairs favoured the female candidate against 4 favouring the male; for positive responses, all 8 discordant pairs favoured the female candidate, with none favouring the male.

The logistic regressions found no significant gender effect on visualisation (Model 1: $\beta = 0.293$, $p = 0.189$) but a significant female advantage in positive responses (Model 3: $\beta = 1.361$, $p = 0.039$), corresponding to an odds ratio of roughly 3.9. Adding geographic controls (Model 4) produced quasi–complete separation and inflated standard errors, so its geographic terms are not interpretable. The exact binomial tests, better suited to the small sample, confirmed a significant female advantage both in visualisation ($p = 0.007$) and in positive responses ($p = 0.008$).

The subgroup analysis localised the response advantage in analytical roles, where all 7 discordant pairs favoured the female candidate ($p = 0.016$), with no differential treatment in technical roles ($p = 1.000$). The bootstrap confirmed the stability of the main result: the positive–response p–value was invariably 0.008 across all 10,000 replications (100% significant at the 1% level), a degenerate but powerful outcome arising because every discordant pair favours the female candidate; the visualisation result was robust but more sensitive to sampling variability (80.5% of replications significant at the 5% level) and should be read with greater caution.

Taken together, the evidence indicates that, at the entry stage of STEM recruitment, identically qualified female graduates are not disadvantaged and in fact receive more favourable recruiter engagement than men. This pattern is consistent with meta–analytic evidence that gender composition predicts the direction of hiring bias and may reflect the tangible impact of diversity initiatives in male–dominated fields (Galos and Coppock, 2023; European Institute for Gender Equality, 2017).

Conclusions

The study makes two contributions. Methodologically, it demonstrates the feasibility and value of conducting correspondence audit experiments within digital labour markets, turning professional networking platforms such as LinkedIn into "natural laboratories" capable of generating timely, large–scale evidence on pressing socio–economic questions (Horton, 2017). Empirically, it provides the first experimental evidence that, at the career–entry stage in the Italian digital labour market, qualified female STEM graduates enjoy equal or higher recruiter responsiveness than identically qualified men, an advantage concentrated in analytical roles.

These findings carry clear policy implications: encouraging girls to enter and persist in STEM education emerges as a productive lever for female employability in high–skilled occupations (Boscaino et al., 2026). The apparent absence of a demand–side penalty at entry relocates the binding constraint upstream in the educational choices and persistence that determine who reaches the STEM labour market, and downstream, in later–career progression. The highest–return policy intervention is therefore investment in the STEM education: early exposure, careers guidance, and sustained support for women's persistence in STEM tertiary tracks and their transition into employment. Our results must, however, be read within the broader landscape of gender inequality. Audit experiments measure direct discrimination but cannot identify its underlying mechanisms (Bertrand and Duflo, 2017), and the results are specific to LinkedIn and constrained by a limited sample. Equity at initial screening does not remove the structural barriers that persist at later career stages, including promotion, pay, leadership representation and the divergence in trajectories around family formation (Goldin, 2014; Blau and Kahn, 2017). Future research should scale up the experiment, extend it to multiple platforms, and incorporate additional covariates. Female STEM participation would be a joint investment in quality education (SDG 4) and gender equality (SDG 5), rather than as the remediation of entry–level hiring discrimination (United Nations, 2015).

Keywords: Gender bias; STEM careers; Correspondence audit; Digital labour market, Educational equity

Tracking Graduate Pathways through Administrative Data: Evidence from Tuscan Universities

Silvia Duranti – Maria Luisa Maitino – Valentina Patacchini

Introduction

In Italy, the university–to–work transition is often considered difficult, as young people tend to experience delayed access to stable employment and heterogeneous returns to tertiary education. However, graduate outcomes cannot be adequately described through a simple employment/non–employment indicator. They require detailed and longitudinal data in order to be properly understood, taking into account the variety of possible academic pathways and, in particular, differences across fields of study. Previous studies on the Italian case have shown the importance of looking not only at employment status, but also at job stability, occupational qualification and the match between degree and job content (Lucisano et al., 2017; Lucisano et al., 2022). A further relevant dimension concerns self–employment, which represents a central pathway of labour–market entry for some disciplinary fields.

In Italy, evidence on graduate employment outcomes is mainly provided by survey–based sources, such as AlmaLaurea and the Istat survey on graduates' transition to work; however, the most recent publicly available Istat microdata refer to earlier cohorts (AlmaLaurea, 2025; Istat, 2015). Administrative data offer a complementary perspective, allowing researchers to follow entire populations of graduates over time and to reconstruct actual employment trajectories on the basis of observed events. Previous administrative evidence on graduates in specific regional contexts focused mainly on dependent employment (Lucisano et al., 2017; Lucisano et al., 2022; Faraoni et al., 2022), while the present contribution extends this perspective by integrating additional information on self–employment, job quality and territorial mobility (Duranti et al., 2026).

This contribution analyses the post–graduation pathways of graduates from Tuscan universities through an integrated longitudinal administrative database. The aim is to show how the integration of university records, mandatory communications on employment relationships and fiscal information can provide a richer picture of university–to–work transitions. The analysis focuses on dependent and para–subordinate employment, self–employment, job quality and territorial mobility among graduates from Tuscan universities. The database covers the 2010–2023 graduation cohorts, with labour–market outcomes observed over specific post–graduation windows.

Data used

The analysis is based on the "Sistema informativo Università toscane", an integrated database developed within a collaboration between Regione Toscana, IRPET, the Universities of Florence, Pisa, Siena, Siena for Foreigners and La Sapienza, with the contribution of the Ministry of Labour (Duranti et al., 2026).

The database covers graduates from Tuscan universities between 2010 and 2023 and combines several administrative sources. University records provide information on graduates' demographic characteristics and academic careers, including type of degree, field of study, year of graduation, age at graduation, final grade and residence. Comunicazioni Obbligatorie record dependent and para–subordinate employment relationships activated throughout Italy, with information on contract type, sector, occupational qualification, place of work, number of employment spells and days worked. Fiscal information provided by Regione Toscana identifies the presence of VAT registration for graduates resident in Tuscany, allowing the analysis to include self–employment.

The database has two main strengths. First, it covers the whole population of graduates from Tuscan universities, rather than a sample or a self–selected group of respondents. Second, it integrates different administrative sources, making it possible to reconstruct longitudinal post–graduation pathways through observed events rather than self–reported information. This represents a complementary approach to survey–based sources such as AlmaLaurea.

The empirical analysis focuses mainly on graduates with master's degrees and single–cycle degrees. Bachelor graduates are excluded from the labour–market analysis because it is not possible to observe whether they continue their studies outside Tuscany. Graduates in Medicine and Surgery are also excluded because their transition to work follows highly specific post–graduate training and professional pathways. Dependent and para–subordinate employment is observed up to five years after graduation. Self–employment is observed through VAT registration within one and three years after graduation and only for graduates resident in Tuscany.

The dataset has some limitations. Self–employment is observed only for graduates resident in Tuscany, because the information is derived from fiscal declarations available to Regione Toscana. The database does not capture work performed abroad, self–employment declared outside Tuscany, or study continuation with master programmes, PhDs or other post–graduate pathways outside Tuscany or abroad.

Methodology

The methodology combines longitudinal descriptive analysis and multivariate modelling.

First, the analysis reconstructs graduates' entry into dependent and para–subordinate employment. The main indicator is the presence of at least one employment relationship within one, three and five years after graduation. This measure is complemented by indicators of job quality and continuity, including number of employment spells, share of days worked, best contract achieved and highest occupational qualification reached. This approach builds on previous analyses of graduate employment based on administrative labour–market data.

Second, the analysis includes self–employment. For graduates resident in Tuscany, fiscal information is used to identify the presence of VAT registration within one and three years after graduation. This allows us to distinguish fields of study in which self–employment is marginal from those in which it represents a structurally relevant pathway of professional insertion. This dimension complements existing evidence on graduates' employment and entrepreneurial outcomes.

Third, dependent work, para–subordinate work and self–employment are combined into a broader measure of positive occupational outcome. For graduates resident in Tuscany, a positive outcome is defined as the presence of at least one dependent or para–subordinate employment relationship, VAT registration, or an already active employment condition at the time of graduation.

A further dimension concerns territorial mobility. By linking graduates' place of residence with the location of the first employment relationship after graduation, the analysis distinguishes between graduates who remain in Tuscany, graduates who leave the region, non–resident graduates who remain in Tuscany after graduation, and non–resident graduates who return to their region of origin or move elsewhere.

Finally, a logistic regression model is used to estimate the probability of a positive occupational outcome within one year after graduation, controlling for field of study, gender, graduation grade, age at graduation and graduation cohort. The model is not intended to provide causal estimates, but to verify whether differences across fields of study persist after accounting for basic individual and academic characteristics.

Results and discussion

The results confirm that university–to–work transitions are highly differentiated across fields of study and cannot be adequately represented by a single employment rate. Considering dependent and para–subordinate work, 58% of graduates have at least one employment relationship within one year after graduation; this share rises to 64% within three years and to 72% within five years. However, access to employment does not necessarily imply stable or high–quality insertion, as also highlighted by previous work on graduate employment, job stability and overeducation.

Some fields show relatively favourable profiles, combining high employment intensity, lower fragmentation and better contractual stability. This is the case, for example, of Economics, Industrial and Information Engineering, and ICT. Other fields display more fragmented trajectories or lower exposure to dependent work. Education and Training is characterised by high employment intensity

but also by high turnover, reflecting the structure of early–career entry into the education sector. Law and Architecture/Civil Engineering show lower levels of dependent employment, but this result requires a broader interpretation because these fields are often associated with professional and self–employed careers.
The inclusion of self–employment modifies the assessment of graduate outcomes. Among master's and single–cycle graduates resident in Tuscany, about 10% have VAT registration within one year after graduation and 19% within three years. The aggregate figure hides strong differences by field of study. Within three years, VAT registration reaches particularly high values in Architecture and Civil Engineering, Agrarian–Forestry and Veterinary studies, Psychology and Law. In these cases, an analysis based only on dependent employment would underestimate the degree of labour–market insertion and could misinterpret professional pathways that are structurally based on self–employment.
Job quality improves over time. In the first phase after graduation, many graduates enter the labour market through fixed–term contracts, traineeships or other non–stable arrangements. Within three and five years, the share of graduates with open–ended contracts or apprenticeships increases, while occupational qualification also improves. A large share of graduates reaches intellectual, managerial or technical occupations, and vertical mismatch appears relatively limited five years after graduation.
Territorial mobility is another relevant dimension. Most Tuscan residents who graduate from Tuscan universities start their dependent or para–subordinate career in Tuscany. The situation is more diversified for non–resident graduates, who are distributed between those who remain in Tuscany, those who return to their region of origin and those who move elsewhere. This highlights the dual role of the Tuscan university system: it attracts students from outside the region, but it does not always retain them in the local labour market after graduation.
The regression analysis confirms the importance of field of study in shaping early occupational outcomes. Differences by gender, graduation cohort, age at graduation and final grade are also observed, but the strongest differences remain associated with disciplinary pathways.

Keywords: University–to–work transition; Administrative longitudinal data; employment outcomes; Self–employment

Beyond Predictive Accuracy: Standardised Tests, School Grades, and University Admission

Giuseppe Caruso – Barbara Luppi – Fabrizio Patriarca

Introduction

The transition from upper secondary school to higher education is a crucial stage in the educational life course. Admission systems determine not only who gains access to university but also how educational opportunities and risks are distributed across students from different social and educational backgrounds. Research on educational stratification has long shown that transitions between educational stages are key points at which inequalities may be reproduced or mitigated, depending on the interaction between institutional rules, prior achievement, and individual resources (Breen & Jonsson, 2005; Lucas, 2001; Jackson, 2013).
University admission systems typically rely on different signals of students' academic preparedness. These include prior school performance, usually measured through high school grades; the educational track attended in upper–secondary school; and standardised admission tests. Standardised tests are generally justified by their ability to provide a common metric across students educated in heterogeneous schools and institutional contexts. However, the empirical literature does not provide unequivocal support for their superiority over school–based measures.

A substantial body of research shows that high school grades often predict medium– and long–term university outcomes better than standardised tests (Geiser & Santelices, 2007; Allensworth & Clark, 2020). One possible explanation is that grades capture not only cognitive skills but also persistence, self–regulation, study habits, and the ability to perform consistently over time (Galla et al., 2019). At the same time, standardised tests may contain additional information and can improve prediction when combined with previous academic records (Kobrin et al., 2008; Sawyer, 2013).

This paper examines the relative and joint predictive value of high school grades, upper–secondary school track, and standardised admission–test scores in Italian higher education. We focus on the TOLC, a computer–based standardised test increasingly used by Italian universities for admission, orientation, and placement purposes. We ask whether the TOLC provides substantial additional predictive information once high school grades and school track are observed; how alternative admission criteria translate into selection errors under different thresholds; and whether these errors are distributed differently across student groups.

The Italian setting is particularly informative because it combines formally broad access to university with substantial educational stratification before university entry. Our central argument is that admission criteria should not be evaluated solely through average measures of predictive performance. Admission decisions are threshold–based and operate under capacity constraints. Therefore, small differences in statistical accuracy may have limited practical consequences, while criteria with similar aggregate predictive performance may distribute false exclusions differently across groups.

Data Used

The analysis uses administrative data from the University of Modena and Reggio Emilia (UNIMORE). The dataset combines information on students' university careers, upper–secondary school achievement, school track, admission–test performance, and selected demographic and socio–economic characteristics.

The analytical sample includes 13,687 first–time entrants to economics and business Bachelor's degree programmes between 2016 and 2021. The analysis is restricted to programmes using the TOLC and to students for whom valid admission–test information is available.

We compare three main admission signals: the total TOLC score, the final upper–secondary school grade, and upper–secondary school track, classified into gymnasium, technical, and professional education. High school grades and total TOLC scores are standardised to facilitate comparison. We also exploit information on individual TOLC components, including mathematics, logic, verbal comprehension, English, science, and knowledge and competencies.

The main outcome is on–time graduation, defined as completion of the Bachelor's degree within the statutory three–year duration. Additional outcomes capture different dimensions of academic success: graduation with honours, enrolment in a Master's degree, dropout, final degree grade, and time to degree.

Fifty–nine per cent of students come from gymnasium tracks, 37 per cent from technical schools, and approximately 4 per cent from professional schools. Thirty–two per cent graduate within the statutory duration, while 19 per cent drop out. Among graduates, the average final degree grade is 98.5 out of 110 and the average time to degree is 3.6 years.

Methodology

The empirical analysis is predictive and decision–oriented rather than causal. Our objective is not to estimate the causal effect of TOLC performance on university outcomes, but to assess the extent to which different information available at admission can discriminate between students with different subsequent academic trajectories.

For binary outcomes, we estimate logit models using alternative combinations of three sets of predictors: TOLC scores, high school grades, and upper–secondary school track. All specifications include degree–programme and enrolment–cohort fixed effects. Cohort fixed effects account for

common temporal factors affecting students entering university in the same year, while programme fixed effects capture systematic differences across degree programmes.

Predictive performance is evaluated using the Area Under the Receiver Operating Characteristic Curve (AUC), a threshold–independent measure of a model's ability to rank students according to subsequent outcomes. We formally compare correlated ROC curves using the non–parametric DeLong test (DeLong et al., 1988).

The analysis decomposes the total TOLC score into its individual components to assess whether particular dimensions of the test account for its predictive content. For continuous outcomes, namely final degree grade and time to degree, we estimate OLS models and compare their explanatory performance.

We assess robustness through three out–of–sample exercises: temporal validation, estimating models on the 2016–2019 cohorts and testing them on the 2020–2021 cohorts; a random 70/30 train–test split; and five–fold cross–validation.

To complement average predictive metrics with a decision–oriented analysis, students are ranked according to predicted probabilities of on–time graduation and increasingly selective admission thresholds are simulated. At each threshold, we compute the share of future successful students who would be incorrectly excluded and the share of future non–on–time students who would be correctly excluded. Finally, we examine the distribution of false exclusions across gender, local versus non–local student status, socio–economic background, and school track.

Results and Discussion

The results show that school–based information provides most of the predictive power available at university entry. For on–time graduation, the TOLC score alone reaches an AUC of 0.723, compared with 0.753 for high school grades and 0.701 for school track. Combining high school grades and school track increases the AUC to 0.771. Adding the TOLC score to this specification further increases the AUC to 0.778.

Thus, the TOLC contains independent predictive information, but its incremental contribution is modest once previous school achievement and educational track are taken into account. The improvement is statistically significant, but amounts to less than one percentage point in AUC terms. This distinction between statistical significance and substantive policy relevance is central to the paper. The component–level analysis reinforces this interpretation. Mathematics is the strongest individual TOLC component, with an AUC of 0.737, and is the only component approaching the performance of the overall test. Logic, verbal comprehension, English, science, and knowledge–and–competencies scores display substantially weaker predictive performance. High school grades outperform each individual test component when estimated on the corresponding samples.

The same ranking is broadly stable across academic outcomes. High school grades outperform TOLC scores in predicting graduation with honours, progression to a Master's degree, and dropout. Combining all three sources of information produces the highest predictive performance, but the incremental contribution of the test remains limited once school–based information is included.

The out–of–sample analyses confirm these findings. In the temporal validation exercise, the AUC is 0.686 for the TOLC–only model and 0.726 for high school grades, while the full model reaches 0.745. The relative ranking is preserved in the random train–test split and five–fold cross–validation.

The decision–oriented analysis shows why differences in predictive accuracy should not be interpreted mechanically. Across simulated exclusion thresholds, admission rules based on high school grades generate fewer false exclusions of future on–time graduates than rules based on TOLC scores, while selection based on school track performs considerably worse.

At the threshold corresponding to the exclusion of both technical and vocational students—41.2 per cent of the sample—school–track selection would incorrectly exclude 32.1 per cent of future on–time graduates. The corresponding figures are 19.9 per cent for TOLC–based selection and 17.3 per cent for grade–based selection. At the same threshold, grades and TOLC scores also correctly exclude larger shares of future non–on–time students than school–track selection.

The heterogeneity analysis further qualifies the aggregate results. High school grades generally produce fewer false exclusions than TOLC scores, but the size of the difference varies across groups and thresholds. Male students experience higher false exclusion rates than female students under both criteria. Differences also emerge between local and non–local students and across socio–economic groups. In particular, students from lower–income backgrounds face higher false exclusion rates under TOLC–based selection than under rules based on school grades.

These findings imply that predictive validity is necessary but not sufficient for evaluating admission criteria. Universities make decisions under capacity constraints, and statistically significant gains in AUC may correspond to small changes in actual selection outcomes. The evidence does not imply that standardised testing is irrelevant. Rather, it suggests reconsidering its institutional role: standardised tests may be more defensible as complementary instruments for orientation, diagnostic support, or additional evaluation where the comparability of prior school performance is particularly uncertain.

Conclusions

This paper compares the predictive and distributive consequences of three types of information available at university admission: high school grades, upper–secondary school track, and standardised admission–test scores.

Three conclusions emerge. First, high school grades are the strongest individual predictor of subsequent university success. School track provides additional information, while the TOLC contains independent but comparatively limited predictive content once school–based indicators are observed.

Second, differences in average predictive performance do not translate mechanically into equivalent differences in admission decisions. Under simulated capacity constraints, grade–based selection generally generates fewer false exclusions than TOLC–based selection, while school–track–based rules perform substantially worse.

Third, admission rules have distributive consequences that aggregate accuracy measures cannot reveal. False exclusions vary across gender, socio–economic background, and student status. Admission systems should therefore be evaluated not only according to how well they predict average academic success, but also according to how they allocate the risks of erroneous exclusion across groups.

The policy implication is not that standardised tests should necessarily be abolished. Rather, their role should be proportionate to their marginal informational contribution. In the setting examined here, the evidence supports using the TOLC primarily as a complementary instrument rather than as a substitute for prior school performance or as a universal primary selection device.

The study has limitations. The data come from a single university, although the institution operates within a framework shared by many Italian public universities. Future research should extend the analysis across institutions and fields of study, particularly where admission is more strongly capacity–constrained or standardised testing plays a stricter formal screening role.

More broadly, university admission systems should be understood as mechanisms of social allocation rather than purely technical prediction exercises. The relevant question is not simply which indicator predicts academic success better on average, but which combination of information improves decisions, for whom, at what cost, and under which institutional constraints.

Keywords: University admission; Standardised tests; Predictive validity; Educational inequality

Grants as Bridge to University Access: The impact of an Information Intervention

Federica Laudisa – Samuele Poy

Introduction

In Italy, student grants are the primary means for removing economic barriers to enable low–income students to access higher education. However, inadequate information among the stakeholders causes a relevant flaw in implementation (Laudisa, 2022). Previous research by IRES Piemonte revealed that a large proportion of first–time tertiary education entrants in Piedmont universities fail to apply for student grants, despite meeting eligibility requirements. We hypothesize that this is due to a lack of awareness of the financial aid opportunities available.

The phenomenon of eligible students failing to apply for financial aid is not unique to Italy, although its magnitude and underlying determinants vary across national contexts. In the United States, students report several reasons for not completing the FAFSA, including the perception that they are not eligible or do not need aid, reluctance to incur debt, the perceived complexity of the application process, and limited knowledge about how to apply (Ifill, 2016). In Germany, misconceptions about eligibility and debt aversion – often linked to overestimating repayment obligations – emerge as key barriers (Riedmiller and Strobl, 2025). Additional obstacles include administrative complexity and co–residence with one's family (Herber and Kalinowski, 2016).

Data used

Our study, conducted on a sample of 6,500 students enrolled in the final year of upper secondary school in Piedmont during 2022/23 school year, identified a substantial information gap regarding university grants. The sample was randomly selected using a stratified sampling design based on two stratification variables: educational track (academic, technical and vocational schools) and geographical area. Overall, approximately 21% of all final–year high school students in the Piedmont region were surveyed.

Most students interviewed lacked basic knowledge of application procedures, eligibility criteria (only 6% were able to accurately identify the eligibility criteria), and even the identity of the grant provider—fewer than 9% correctly identified EDISU Piemonte as the awarding body, while 72% had never even heard of it. We also found a significant correlation between knowledge of grants and individual characteristics, including gender, grade point average, school track, family background, and participation in guidance activities.

Female students, students with parents who have tertiary education, those with higher grade point averages, and students following academic tracks—considered the most advantaged—are more informed and significantly more likely to aspire to further education. This finding is not particularly surprising (INVALSI, 2025; AlmaDiploma, 2025). However, this study also shows that, all else being equal, more active participation in guidance activities and greater awareness of university grants and tuition fee exemptions significantly increase the likelihood of intending to enroll in university. This suggests that guidance activities and student support may play a crucial role in mitigating the influence of family background in socioeconomically disadvantaged contexts. To test this hypothesis, we introduced a new research question: Does informing students about financial aid influence university enrolment? To address this question, we conducted a randomized controlled trial.

Methodology

Using a randomized controlled trial, we examined the impact of an informational session about financial aid on university enrollment and grant application rates. In other words, we investigated whether raising awareness of financial aid opportunities influences students' choices to pursue higher education and, if they do enroll, whether it increases the number of grant applications submitted.

Participating students were randomized into treatment and control groups. The treatment group attended a 30–minute talk providing them with detailed information about the grant, including eligibility requirements, application procedures, and award amounts. Conversely, the control group received no such information. Follow–up interviews were conducted about 6 to 7 months after students graduated from secondary school to determine how many had enrolled in university and/or applied for a grant. By comparing the rates of the treatment and control groups, we were able to evaluate the impact of having attended the informational session.

Results and discussion

This randomized controlled trial contributes to a relatively sparse strand of literature. At the national level, Abbiati *et al.* (2017) conducted a preliminary study where upper secondary students attended three separate informational sessions covering various aspects such as costs, employment returns, drop–out rates by degree course, and scholarships. The aim was to ascertain whether these guidance activities could influence enrollment rates and alleviate the impact of social background. Although no impact was observed, at least regarding enrollment rates, the international literature shows different but not incontrovertible outcomes. Overall, the literature tends to suggest that the impact is most likely when informational sessions are coupled with personalized counseling activities (Herbaut & Geven, 2020). The only exception is a study by Loyalka *et al*. (2013), which shows that a single informational meeting increased university enrolment.

Similarly, our study revealed that a one–time intervention on financial aid had a positive and significant effect on enrollment rates, estimated at +2.8 percentage points. Certain subgroups saw an even larger impact: female students, whose enrollment increased by 4 percentage points compared to the control group; students with limited participation in guidance activities (i.e., the least informed); and, in particular, students who were undecided due to financial concerns, among whom enrollment increased by 8.3 percentage points, showing that financial aid awareness plays a decisive role in their decision–making. This suggests that grants can make the difference in choosing whether to enroll in university, particularly when economic concerns contribute to indecision.

However, the effect was null for vocational track students, suggesting that a more intensive information program may have been necessary since their training is specifically geared towards entering the labor market. This hypothesis finds support in research by Martini *et al.* (2021), who observed the impact of the “Percorsi” program on the probability of university enrollment among vocational school students.

Informing students about grant opportunities not only influences college enrollment but also significantly boosts the number of grant applications submitted (+7.4 p.p). This impact is even more pronounced (+15.9 p.p.) among students hailing from lower socioeconomic family backgrounds, who are the main beneficiaries of financial aid programs.

Conclusions

In conclusion, our findings underscore the importance of providing information about financial aid to students as a strategic approach to increasing access to higher education in Italy. Our recommendation is that the Ministry systematically disseminate information through digital platforms and school applications accessible to both students and parents — an intervention that is virtually cost–free.

However, information alone does not appear to be sufficient for vocational students or for those who remain undecided for reasons unrelated to financial constraints. For these groups, more personalized guidance and complementary support services—such as tutoring, psychological counselling, or mentoring—may be required. This interpretation is supported by students’ own responses: among those who were undecided about pursuing higher education, 16% cited concerns such as the perceived difficulty of university studies, uncertainty about their academic abilities, or unclear ideas about which course to undertake.

In fact, when we asked students why they were undecided about continuing their studies, 16% of them wrote their own reasons. These responses point to perceived difficulty of university studies,

anxiety or uncertainty about their abilities, or unclear ideas about which course to undertake, showing why this type of effort is needed. There is also substantial scope for improving the implementation of these forms of support.

Keywords: Information; Financial aid; Randomized controlled trial; School–University transition

SESSION 9. DIGITAL COMPETENCE AND GENERATIVE AI IN EDUCATION: NEW CHALLENGES FOR TEACHING AND ASSESSMENT

ORGANIZERS: INVALSI – UNIVERSITY OF BERGAMO

COORDINATORS: PAOLO BARABANTI – MARCO GIGANTI

5TH NOVEMBER: 8.30 A.M. – 10.30 A.M. {AULA B0.1 – RESEARCH 9}

Digital Competences among First– and Second–Generation Migrant Students in Italy: Evidence from INVALSI Data

Valentina Ferri – Giovanna Filosa – Patrizia Rinaldi

Introduction

Digital competence has become a key dimension of educational equity, social participation, and future labour market inclusion. It can no longer be understood solely as access to devices, connectivity, or digital infrastructure. Rather, it refers to a multidimensional set of abilities that includes searching, evaluating and managing information, communicating and collaborating in digital environments, creating digital content, and using technologies safely and responsibly. This interpretation is consistent with the European DigComp 2.2 framework, which defines digital competence as the combination of knowledge, skills and attitudes needed to engage with digital technologies in a confident, critical and safe way (Vuorikari, Kluzer, & Punie, 2022).

The theoretical background of this paper draws on the literature on digital inequality and the digital divide. Early studies mainly focused on access to ICT, while more recent research has shifted attention to differences in digital skills, types of use and the benefits individuals derive from technology. Hargittai's concept of the second–level digital divide is particularly relevant, as it shows that access alone does not eliminate inequality: individuals differ in their capacity to use digital tools effectively, and these differences are socially structured (Hargittai, 2002). Van Dijk further conceptualises digital inequality as a cumulative and dynamic process involving motivation, physical access, skills, usage and outcomes, all of which interact with pre–existing social inequalities such as socio–economic status, education, gender and territorial resources (van Dijk, 2020).

This framework is especially relevant when analysing students with a migrant background. International research shows that migrant–origin students may experience educational disadvantage linked to socio–economic background, language barriers, school segregation and unequal access to learning opportunities. OECD analyses based on PISA data distinguish between first–generation immigrant students, who were born abroad and whose parents were also born abroad, and second–generation immigrant students, who were born in the country of assessment but whose parents were born abroad (OECD, 2023). Recent evidence from ICILS 2023 also highlights substantial differences in digital literacy by socio–economic and migrant background, confirming that digital competences should be analysed as part of broader educational inequalities (International Association for the Evaluation of Educational Achievement [IEA], 2025).

Against this background, the paper investigates inequalities in digital competences among Grade 10 students in Italy, with a specific focus on first– and second–generation migrant students. The study aims to identify the main individual, family, school–track and territorial factors associated with the school and territorial digital divide. More specifically, it examines the role of gender, socio–economic and cultural background, migrant origin, geographical area and type of upper–secondary school. The analysis also assesses whether differences associated with migrant origin persist after controlling for socio–economic background and other relevant student and school–context characteristics.

Data Used

The analysis will use the INVALSI Grade 10 sample dataset on digital competences for the school year 2024/25. The record layout includes anonymous identifiers for schools, classes and students, as well as student, class and school weights, which allow the analysis to account for the survey structure (INVALSI, 2024/25).

The dataset contains information on several student characteristics: gender, month and year of birth, place of birth, regularity in schooling, migrant origin, father's and mother's place of birth, school timetable, school type, region and macro–geographical area (INVALSI, 2024/25). The key independent variable is migrant origin, which distinguishes native students, first–generation foreign–background students and second–generation foreign–background students. The main socio–economic control is the ESCS student index, which measures the students' socio–economic and cultural background.

School–track differences will be analysed through the variable identifying the type of upper–secondary school: classical, scientific and linguistic lyceums; other lyceums; technical institutes; vocational institutes; and missing information. Territorial differences will be examined using both region codes and five macro–areas: North–West, North–East, Centre, South, and South and Islands (INVALSI, 2024/25).

Digital competence outcomes are measured through WLE scores based on Rasch modelling and reported on a national scale. The dataset includes four domains: Information and Data, Communication and Collaboration, Digital Content Creation, and Safety. For each domain, proficiency levels are also available and classified as basic, intermediate and advanced (INVALSI, 2024/25). This structure makes it possible to analyse both continuous differences in digital competence scores and categorical differences in the probability of reaching higher levels of competence.

Methodology

The empirical strategy will proceed in three main steps. First, the study will provide a descriptive analysis of digital competence inequalities by migrant origin, gender, ESCS, geographical area and school type. Weighted means and distributions across proficiency levels will be reported for the four digital domains. This stage will allow us to assess whether inequalities are concentrated in specific areas of competence or whether they appear consistently across the whole digital competence profile.

Second, the paper will estimate multivariate statistical and econometric models to evaluate the association between student background and digital competence. Linear regression models will be estimated using the WLE scores as dependent variables. Separate models will be run for each domain: Information and Data, Communication and Collaboration, Digital Content Creation, and Safety. The baseline specification will include migrant origin, gender, ESCS, school regularity, school type, macro–area and region controls. This approach is consistent with the literature showing that digital inequalities are embedded in broader social and educational inequalities rather than being reducible to individual access to technology alone (Hargittai, 2002; van Dijk, 2020).

Additional specifications will include interaction terms between migrant origin and ESCS, migrant origin and school type, and migrant origin and geographical area. These models will assess whether the digital competence gap varies according to socio–economic resources, institutional pathways and territorial contexts. This is particularly relevant because previous research suggests that ICT–related educational inequalities may interact with migrant background and socio–economic disadvantage, rather than affecting all students in the same way (Rodrigues, 2018; Rodrigues & Biagi, 2017).

Third, categorical models will be estimated using proficiency levels as dependent variables. Ordered logit or multinomial models will analyse the probability of being placed at the basic, intermediate or advanced level in each domain. This step is important from a policy perspective because digital vulnerability is not only reflected in average score differences but also in the concentration of students at the lowest levels of competence.

As an exploratory extension, the analysis may apply clustering techniques to identify differentiated student profiles. These profiles may include students with high digital competencies across domains, students with good disciplinary outcomes but digital fragilities, and students who appear vulnerable across multiple dimensions. This typological approach would complement the regression models by identifying groups of students for whom targeted interventions may be especially relevant.

Results and Discussion

The analysis will quantify the extent to which digital competence gaps are associated with migrant origin and whether such differences are mediated by socio–economic background, school track and territory. The central research question is whether first– and second–generation migrant students differ from native students in their levels of digital competence and whether these differences remain statistically significant after accounting for ESCS and other individual and contextual variables.

The paper will contribute to three main areas of discussion. First, it will assess whether migrant–background students should be considered as a homogeneous group or whether first– and second–generation students display distinct digital competence profiles. This distinction is analytically important because first–generation students may face barriers linked to migration history, language acquisition and school integration, while second–generation students may experience different forms of disadvantage related to family background, social capital and educational trajectories. The OECD distinction between first– and second–generation immigrant students provides a useful comparative framework for interpreting these differences (OECD, 2023).

Second, the study will examine the role of socio–economic and cultural background. If the inclusion of ESCS substantially reduces migrant–origin differences, the digital divide should be interpreted mainly as part of broader social inequality. If, instead, migrant–origin gaps persist after controlling for ESCS, the findings would suggest the presence of additional mechanisms, such as linguistic barriers, unequal access to informal learning opportunities, differences in parental support or varying school–level resources. This interpretation is consistent with research showing that digital inequalities are not only technological but also social, cultural and educational (Hargittai, 2002; van Dijk, 2020).

Third, the analysis will focus on the territorial and institutional dimensions of the digital divide. Differences across regions, macro–areas and school types may indicate that digital competences are shaped by unequal educational opportunities, curricular pathways and local resources. In this sense, digital competence can be understood not only as an individual outcome but also as an indicator of the opportunities provided by schools and territories. International evidence from ICILS confirms that digital literacy differs substantially across education systems and social groups, including by socio–economic and migrant background (IEA, 2025).

The discussion will interpret the results in light of the second–level digital divide theory. The paper argues that schools should not treat digital competence as a purely individual attribute or as a spontaneous consequence of young people's exposure to technology. Instead, digital competence develops through the interaction between family resources, school opportunities, peer environments and territorial infrastructures. This interpretation supports the view that digital inequalities may reinforce pre–existing educational inequalities if they are not addressed through explicit and targeted policy action (Hargittai, 2002; van Dijk, 2020).

Conclusions

This paper will provide new evidence on digital competence inequalities among Grade 10 students in Italy, with a specific focus on first– and second–generation migrant students. By using INVALSI data on digital competences, the study will move beyond general measures of school achievement and examine a set of skills that are increasingly central to educational success, social inclusion and participation in technological and labour–market transformations.

The expected contribution is twofold. From a scientific perspective, the paper connects research on digital inequality with the literature on migrant–background students and educational stratification. From a policy perspective, it offers evidence to support targeted interventions aimed at reducing

digital gaps across student groups, territories and school tracks. The results may help identify where support is most needed: among students with low socio–economic background, in specific geographical areas, in particular school pathways, or among first– and second–generation students facing cumulative disadvantages.
Overall, the study frames digital competence as both an educational outcome and an indicator of opportunity. Reducing digital competence gaps should therefore be considered part **of a broader strategy to promote equity, inclusion and preparedness for technological change.**

Keywords: Digital Competences; Migrant Students; INVALSI Data; Digital Divide

Reading Achievement, Socioeconomic Inequality, and Adolescents' Resistance to Misinformation Globally

Mobarak Hossain – Jason Jabbari

Introduction

Adolescents in the contemporary world navigate an unprecedented and overwhelming amount of digital information and misinformation at a time when their minds are still developing (Crone and Konijn 2018; McGrew et al. 2018). While past generations have also faced challenges to democratic norms, the current speed and reach of online misinformation pose new risks and threats to young people as they approach adulthood, political participation, and major life transitions (Zakaria 2024). Misinformation and viral content on social media are often shared by superficially looking at the headline, without scrutinizing the content (Sundar et al. 2025).
This paper examines adolescents' resistance to misinformation across 49 countries. We conceptualize misinformation resistance as a set of observable practices that reflect students' engagement with, evaluation of, and sharing of online information. Schools remain important for nurturing the critical skills of students and the civic dispositions necessary for informed participation in a democratic society (Dewey 1915; Gutmann 1987). However, evidence suggests that students struggle with evaluating basic digital sources and often rely on weak signs of credibility, such as a website's appearance (Breakstone et al. 2021; Caulfield and Wineburg 2023).
Prior research shows that literacy is a key predictor of students' capacity to locate, assess, and share information, and that higher achievement can act as a protective mechanism against misinformation (Hahnel et al. 2018; Kahne and Bowyer 2017). At the same time, misinformation resistance is not shaped by achievement alone. Cognitive heuristics, confirmation bias, motivated reasoning, and broader social contexts also influence susceptibility to misinformation (Flynn, Nyhan, and Reifler 2017; van der Linden 2023). Socioeconomic status may therefore moderate the relationship between academic skills and misinformation resistance, because students from lower–SES backgrounds may have fewer opportunities to build literacy, critical–thinking skills, and epistemic strategies outside school (Dolean et al. 2021; Hargittai 2010; Livingstone and Helsper 2007). We address three questions: how misinformation resistance varies globally, how reading achievement is associated with misinformation resistance, and how SES interacts with reading achievement in predicting misinformation resistance.

Data used

We use data from the Program for International Student Assessment (PISA) 2022 (OECD n.d.). The analytic sample includes 232,901 students in 11,221 schools across 49 countries for which the dataset includes surveyed indicators on misinformation resistance. PISA is especially useful because it provides cross–national information on 15–year–olds' academic skills, family background, school context, and digital information practices. This age group is particularly important because adolescents are exposed frequently to social media platforms and online content, while also being in

a formative period for civic and epistemic development (Brown and Gummerum 2025; Wang and Togher 2024).

Methodology

The dependent variable is a misinformation resistance index based on eight PISA 2022 items. These items capture how often students report engaging in critical online behaviors, including comparing sources, checking accuracy before sharing information, discussing online content with teachers, friends, or family members, avoiding sharing made–up information without flagging it, and flagging wrong information when they encounter it online. Two items, trusting what one reads online and sharing made–up information without flagging its inaccuracy, are reverse–coded so that higher values indicate more careful or skeptical engagement with online content. Internal consistency analysis shows good reliability, supporting the construction of a composite index.

The key predictor is reading achievement, measured using PISA plausible values. Reading achievement captures students' ability to understand, evaluate, and reflect on written texts across formats and contexts, which we theoretically link to the capacity to critically assess online information. Socioeconomic status is measured using the PISA economic, social, and cultural status index and categorized into within–country quintiles. We also control for sex, immigration background, homework time in reading, ICT access at home and school, student–teacher ratio, school urbanity, GDP per capita, inequality, and political regime characteristics.

We use multilevel models with students nested within schools and countries. This allows us to estimate the association between reading achievement and misinformation resistance while accounting for school– and country–level variation. We then introduce interactions between SES quintiles and reading achievement to examine whether the role of reading achievement differs by socioeconomic background. Robustness checks use mathematics and science achievement and an alternative misinformation resistance index based on principal component analysis.

Results and discussion

The descriptive results show that many adolescents report basic misinformation–checking behaviors. Around 70 percent of students report comparing sources and checking accuracy before sharing information. However, misinformation resistance is far from universal. A large proportion of students still trust what they read online, do not flag wrong information, and do not discuss misinformation with parents, friends, teachers, or classmates. These patterns suggest that adolescents are not simply passive consumers of online information, but many remain vulnerable to misinformation sharing and its consequences.

The results also show a consistent positive SES gradient in misinformation resistance. Higher–SES students are more likely to report careful and skeptical engagement with online information than lower–SES students. This pattern is consistent with prior research suggesting that social class shapes digital information environments and opportunities to develop critical digital skills (Lindell and Mikkelsen Båge 2023). It also indicates that misinformation resistance is an inequality issue, not only a digital literacy issue.

Reading achievement is positively and significantly associated with misinformation resistance across all models. The association remains after accounting for SES, student background, school context, household and school ICT access, urbanity, and country–level characteristics. This suggests that stronger reading skills are associated with higher levels of skepticism toward online misinformation. The finding is consistent with earlier evidence that literacy skills underpin students' ability to find, evaluate, and interpret online information (Hahnel et al. 2018; Kahne and Bowyer 2017).

The interaction between SES and reading achievement provides the main contribution of the paper. Higher–SES students report higher misinformation resistance, but the positive association between reading achievement and misinformation resistance is stronger among lower–SES students. In other words, disadvantaged students start from a lower baseline, but gains in reading achievement are associated with larger improvements in misinformation resistance. This is consistent with a compensatory model of skill formation. Among lower–SES students, reading proficiency may

partially offset the absence of other familial and communal resources, including parental guidance, access to diverse information sources, and structured opportunities to discuss current events. Among higher–SES students, the association is weaker because supportive contexts may already facilitate critical engagement with information (Bourdieu 1986; Lareau 2011).
The results also show that most variation in misinformation resistance lies at the individual level rather than the school or country level. National context, including political, economic, and cultural factors, explains smaller variation once individual and school characteristics are taken into account. Robustness checks using mathematics and science achievement produce substantively similar results, suggesting that broader analytical and problem–solving skills also relate to misinformation resistance. The PCA–based misinformation index also produces consistent findings.

Conclusions

This study offers one of the first large–scale, cross–national analyses of adolescents' resistance to misinformation using PISA 2022 data. The findings show that many students engage in basic verification practices, but substantial proportions remain vulnerable to trusting, sharing, or failing to challenge misinformation. Reading achievement is positively associated with misinformation resistance, but this relationship is socially patterned. SES differences are large, yet achievement gains are especially important for disadvantaged students.
The study therefore shows that combating misinformation requires more than digital access. It requires cognitive empowerment and equity. Investments in literacy, information literacy, and targeted support for disadvantaged students can yield disproportionate gains in critical digital resilience. At the same time, literacy skills alone cannot fully offset structural inequalities in access, exposure, and support. Misinformation resistance is shaped by individual skills and the social contexts in which those skills are developed and used.

Keywords: Misinformation; Adolescence; Inequality; Reading achievement; PISA

Towards Automated Classification of Open–Ended Responses: A Preliminary Study Using Transformer–Based Language Models

Michele Marsili – Fiammetta Noccioli – Emiliano Campodifiori – Marcello Napoli

Introduction

With the digitization of educational assessments and the transition to computer–based testing (CBT), automated approaches to response scoring have become increasingly feasible. Recent advances in machine learning have also influenced the field of open–ended question responses through the development of Large Language Models (LLMs). Broadly defined as deep learning algorithms built entirely on the Transformer architecture introduced by Vaswani et al. (2017) and trained on large–scale text datasets to comprehend and generate human language, these models have significantly improved natural language understanding. Consequently, LLMs are particularly suitable for educational assessment tasks that require the interpretation of complex written responses, effectively redefining the boundaries of natural language processing (NLP). Several international studies have been conducted on this aspect in the context of large–scale assessments where a considerable number of open–ended responses are collected. The literature shows a progressive evolution from rule–based approaches to advanced deep learning systems. Yamamoto et al. (2017) proposed a machine–supported coding framework based on exact matching of previously scored responses, while Okubo et al. (2023) demonstrated the feasibility of multilingual deep learning models for automatic scoring in PISA. More recently, Capdehourat et al. (2025) investigated the

effectiveness of LLMs for grading open–ended responses, emphasizing the role of prompt engineering. Jauhiainen and Guerra (2026) focused on the language's issues testing GPT–4o. As noted by Zhu et al. (2025), several challenges are inherent to the automated grading process. Student answers often contain irrelevant or repetitive information, making it difficult to isolate the key points necessary for accurate grading. Furthermore, responses that differ significantly from the reference answers, in terms of length or wording, may still be conceptually correct and relevant to the question. The aim of this preliminary study is to investigate whether Transformer–based LLMs can accurately grade open–ended responses in Italian language, achieving results comparable to human evaluation. It also compares different LLM–based approaches to identify the models that offer the best balance between grading accuracy and computational efficiency. More specifically, the study seeks to determine whether current LLMs can achieve a level of performance that would make them suitable for future operational use in large–scale assessments. If successful, these advancements could significantly improve assessment processes by reducing scoring time and operational costs while maintaining a high standard of evaluation quality.

Data

This study utilizes data from standardized assessments administered across Italian schools by INVALSI. Specifically, we selected a set of open–ended student responses from the Grade 13 Mathematics assessment during the 2025–2026 school year. The analysis focuses on highly discursive items characterized by varying levels of difficulty. Consequently, the dataset includes the students' written text alongside their binary human–graded scores (correct/incorrect). Non–responses were excluded from the sample. For each item, two experimental scenarios were designed: the first utilizes a restricted dataset of approximately 1,500 human–graded responses, while the second expands the scale to a larger dataset of roughly 20,000 responses. This setup allows us to evaluate whether, and to what extent, the training dataset size influences the performance of each experimented model.

Methodology

Student responses are analyzed from a quantitative perspective using a combination of Transformer–based language models and traditional machine learning algorithms. The study is organized into several phases. First, data cleaning and preprocessing procedures are applied to ensure data quality and consistency by correcting misspellings, removing anomalies, and standardizing text formatting (e.g., capitalization). The Transformer architecture (Vaswani et al., 2017) represents the foundation of most contemporary natural language processing systems, including encoder–based models such as BERT and decoder–based models such as GPT. For the Transformer–based approach, models belonging to the BERT family (Devlin et al., 2019; Liu et al., 2019; Reimers & Gurevych, 2019), RoBERTa and Sentence–BERT, including multilingual variants, are employed. BERT (Bidirectional Encoder Representations from Transformers) captures contextual information bidirectionally, enabling a deep understanding of textual meaning. These models are used to generate text embeddings, namely numerical vector representations that encode the semantic and contextual information contained in student responses. The generated embeddings can either be directly used for classification through fine–tuning or serve as input features for traditional machine learning classifiers, such as Support Vector Machines (SVM) and XGBoost (Xiao et al., 2025). In particular, several machine learning models, suitable for classification problems and scaled well to large dataset suitable, were evaluated to assess their ability to identify the correct category of student responses. To ensure a representative distribution of response categories, stratified sampling is adopted during the train–test split. In addition, cross–validation and hyperparameter optimization are performed to improve model performance and robustness.Model effectiveness is assessed by using standard classification metrics, including accuracy, precision, recall, and F1–score. Finally, model predictions are compared with human coding outcomes to examine the level of agreement and identify instances of mismatch between automated and manual classifications. This

comparison provides insights into the strengths and limitations of the different approaches in categorizing student responses.

Results and discussion

In the first scenario, RoBERTa combined with SVM showed the most stable performance across the different items, maintaining accuracy values between 0.84 and 0.87, corresponding to an agreement rate with human coding outcomes, and also a precision between 0.80 and 0.85.

Although the RoBERTa–XGBoost model achieved the highest overall accuracy (0.88) on two items, its performance was less consistent across the full set of items.

A similar pattern emerged in the second scenario. The RoBERTa–SVM model again proved to be the most consistent approach, achieving accuracy values ranging from 0.84 to 0.86 across the different items. In contrast, RoBERTa–XGBoost attained the highest performance on specific items, reaching an accuracy of 0.91, although with greater variability across items.

Similar trends were observed for precision, recall and F1–score, indicating that model performance was not limited to a single evaluation metric.

The distribution of response categories reflects the actual difficulty of the items. Easier items naturally generated a higher proportion of correct responses, resulting in class imbalance. Rather than artificially balancing the data, the study retained the original distributions to better represent operational assessment conditions. The models maintained satisfactory performance despite differences in class distributions across items, suggesting robustness to variations in item difficulty. Although F1–scores were generally lower than accuracy values, the differences remained limited, indicating a relatively balanced performance across response categories. It will be important to evaluate balancing strategies without compromising the representativeness of operational testing conditions.

As expected, the availability of larger labeled datasets led to improvements in classification performance across most models. However, a particularly relevant finding is that high accuracy and precision levels were already achieved in the first scenario, suggesting that Transformer–based embeddings combined with traditional machine learning classifiers can provide reliable results even when trained on more limited amounts of labeled data.

Conclusions

Among the tested methods, RoBERTa models achieved the highest classification performance, demonstrating that AI–based approaches can support the evaluation of open–ended responses from the quantitative perspective.

Although the proposed models achieved promising results, they should be viewed as decision–support tools rather than replacements for human evaluation. Human judgment remains essential for handling ambiguous, complex and context–dependent responses, as well as for ensuring the validity and fairness of the assessment process. Future work should investigate operational human in the loop procedures, while also exploring the generalizability of the models across different subjects, item formats and question types.

Finally, despite the preliminary nature of the present study, the adoption of AI–based systems in educational assessment should be accompanied by careful consideration of privacy, transparency, and accountability issues.

Keywords: Open–Ended Responses; Educational Assessment; Transformer–Based Models; Machine Learning

School Digital Readiness and Territorial Inequalities in Secondary Education: A Document–Based Mapping of Schools

Ilaria Ravasi

Theoretical Framework

Digital competences are at the top of the European policy agenda, with the aim of strengthening them to support the ongoing digital transformation. The *Digital Compass 2030* (European Commission, 2020) and the *European Pillar of Social Rights Action Plan* (European Commission, 2021) set the target of ensuring that at least 80% of the population possesses basic digital skills by 2030. In Italy, in alignment with the enabling structural reforms foreseen by the National Recovery and Resilience Plan (PNRR, Mission 4)[5], a process has been initiated to establish a national system for the assessment of digital competences starting from the 2024–2025 school year.

The increasing diffusion of Generative Artificial Intelligence technologies in educational contexts has further expanded the relevance of digital competences, highlighting that they cannot be reduced to a purely technical skill set. At the same time, this development is raising new questions concerning learning assessment, educational equity, and the transformation of teaching practices. In line with the European framework *DigComp 2.2* and its recent update *DigComp 3.0* (European Commission, 2025), digital competence is understood as encompassing cognitive, critical, ethical, and reflective dimensions, including the ability to understand, evaluate, and use emerging technologies in an informed and responsible manner. Among these, AI systems are increasingly reshaping teaching, learning, and assessment practices.

The first large–scale sample–based assessment of digital competences at Grade 10 (INVALSI, 2025) has produced results that warrant careful attention. When disaggregated by academic track study across a three–level scale (low, intermediate, advanced), a higher concentration of students at the intermediate level is observed in classical, scientific, and linguistic upper secondary schools. In other lyceum tracks and technical institutes, results appear similar or slightly lower, while less favourable outcomes are observed in vocational schools.

Although these findings must be interpreted cautiously due to the absence of longitudinal trend data, they nevertheless raise important considerations regarding educational tracking (Hanushek & Woessmann, 2014) and educational inequalities. The literature has shown that differentiated educational pathways tend to reproduce social inequalities, as school choices are strongly associated with students' socio–economic and cultural backgrounds, territorial characteristics, and teachers' guidance recommendations (Ballarino, Panichella & Triventi, 2014; Polesel & Leahy, 2019). In the specific case of digital competences, tracking may generate an educational digital divide: lyceum tracks tend to promote a more critical and analytical use of technologies, whereas vocational pathways are more likely to emphasise technical and instrumental uses, thereby structuring differentiated levels of digital competence (Gui & Argentin, 2011).

Adopting an ecological perspective (Bronfenbrenner, 2007), the present study develops a territorial mapping as a preliminary and exploratory phase. The aim is to investigate how different school and territorial configurations contribute to shaping the conditions for the development of students' digital competences. To this end, a mapping of upper secondary schools in the province of Bergamo is conducted, with the purpose of identifying distinct profiles of school digital readiness and the organisational strategies through which schools promote the development of digital competences.

In this perspective, school digital readiness is also conceptualised as the capacity of educational institutions to create organisational and cultural conditions that foster the critical integration of emerging digital technologies. Within this framework, the presence of references to artificial intelligence in school documents is considered an exploratory dimension of digital readiness, useful

[5] https://pnrr.istruzione.it/wp-content/uploads/2022/07/PIANO_SCUOLA_4.0_VERSIONE_GRAFICA.pdf

for capturing early signals of organisational transformation. This mapping constitutes an interpretative tool for examining differences observed in national assessments of digital competences, providing a contextual framework through which variability across educational tracks and schools can be understood.

Method

The study develops a territorial mapping and represents the exploratory phase of a broader sequential mixed–method research design, structured into three interrelated stages, aimed at investigating the relationship between digital competences and learning vulnerabilities. The selected study area is the province of Bergamo. This choice is primarily grounded in criteria of scientific relevance: the Bergamo context is characterised by marked socio–economic and territorial heterogeneity (ranging from urban centres to mountainous and peripheral areas) as well as by a highly diversified structure of upper secondary education provision. This allows for the observation of different configurations of educational pathways, territorial contexts, and educational opportunities. In addition, the selection was informed by contingent and institutional considerations related to the possibility of collaborating with local institutional stakeholders.

The study involves all 39–state upper secondary schools in the province. Private schools were excluded to ensure homogeneity and full comparability of the documentary corpus. The dataset consists of territorial indicators contained in official provincial institutional reports and school strategic documents, namely the *PTOF*, the *RAV*, and the *Social Reporting* documents. This documentary corpus is intended to ensure a standardised and institutionally grounded information base with respect to four analytical dimensions structuring the mapping: (1) socio–territorial context, (2) school profile and field of study, (3) school digital readiness, and (4) improvement priorities. These dimensions are identified based on the relevant literature.

Within the dimension of school digital readiness, the potential presence of references to artificial intelligence and emerging technologies in strategic school documents is also explored. The documents are subjected to qualitative content analysis aimed at constructing a typology of schools and identifying criteria for the subsequent analysis of the relationship between digital competences and learning vulnerabilities using INVALSI data.

Discussion

In line with the relevant scientific literature, the working hypothesis is that different territorial and organisational configurations of schools contribute to explaining part of the variability observed in the development of digital competences across different fields of study, suggesting an unequal distribution of educational opportunities across educational pathways and local contexts. These differences may also translate into unequal opportunities for access to, and integration of, generative AI technologies, which are conceptualised as an indicator of schools' organisational capacity to integrate emerging technologies into educational processes, both in terms of competence development and in terms of the establishment of institutional conditions conducive to AI–supported teaching and assessment practices.

The empirical analysis is currently ongoing; the results will be presented and discussed at the seminar. From a theoretical perspective, the study contributes to the integration of three main strands of research: educational tracking, digital competences, and learning vulnerabilities. Adopting an ecological perspective, the research conceptualises schools as differentiated educational ecosystems in which institutional, social, and territorial factors jointly shape learning opportunities. The outcomes of the mapping will provide a reference framework for subsequent empirical analyses based on INVALSI assessment data in italian, mathematics, and digital competences, with the aim of achieving a deeper understanding of the strengths and weaknesses of the identified school institutions.

Conclusions

The study proposes an interpretative model of the contexts in which digital competences develop within upper secondary schools, using territorial mapping as an analytical tool to examine different

school configurations. Focusing on the province of Bergamo, the research adopts a perspective aimed at highlighting the role of local contexts and educational pathways in shaping differentiated learning opportunities.
By integrating the literature on educational tracking, digital inequalities, and learning vulnerabilities, the study develops an ecological perspective in which schools are conceptualised as differentiated educational ecosystems that mediate access to digital opportunities and influence students' educational trajectories.
From a methodological standpoint, the study develops a multidimensional analytical framework based on documentary sources (*RAV*, *PTOF*, *Social Reporting*) and contextual indicators, offering a replicable approach for mapping school systems beyond individual performance measures alone. The proposed framework enables the identification of school profiles characterised by different configurations of territorial context, tracking structure, levels of school digital readiness, and strategic priorities.
Overall, the proposed mapping represents a preliminary but necessary phase in the development of a research design aimed at exploring the relationships between digital competences, educational tracking, and learning vulnerabilities an area that remains insufficiently investigated at a systematic territorial and institutional level.
Furthermore, in line with evidence from INVALSI assessments and the orientations of major European and national education policies, the study underscores the importance of in–depth analysis of specific territorial contexts to understand how different school and territorial configurations contribute to either an equitable or uneven distribution of educational opportunities. In this perspective, the proposed mapping may help to elucidate how territorial and school–level differences influence the emergence of enabling or constraining conditions for the integration of generative artificial intelligence technologies into educational processes, with potential implications for educational inequalities.

Keywords: Digital competences; School digital readiness; School mapping; Educational inequalities

Open Digital Badges: A Preliminary Analysis on the INVALSI Case Study

Veronica Pastori – Antonio Severoni

Introduction

Open Digital Badges (ODBs) were introduced in 2012 by the Mozilla Foundation with the initial purpose of recognising and sharing learning achievements in informal educational contexts. Over time, their use has expanded to formal education systems, including primary and secondary schools, higher education institutions, and professional training programs (Young et al., 2019).
The creation of the Open Badge Infrastructure (OBI) played a key role in this expansion, providing a shared technical standard ensuring interoperability across platforms and enabling badges to function as verifiable, portable credentials (Casilli & Hickey, 2016). Each badge is associated with metadata that documents the issuing organisation, the criteria required to earn it, and evidence of the competencies attained (Tharindu et al., 2017). This structure distinguishes ODBs from simple digital certificates, providing a more reliable and transparent record of achievement.
ODBs serve several core functions: they act as a source of intrinsic and extrinsic motivation (Abramovich et al., 2013; Denny, 2013; Roy & Clark, 2019), promote self–regulated learning and learner autonomy (Randall et al., 2013), and support the contextualisation of skills, contributing to building public trust in individual accomplishments (Casilli & Hickey, 2016). According to Antin and Churchill (2011), in digital contexts, ODBs reinforce five key socio–psychological functions: goal setting, instruction, reputation, status/affirmation, and group identification.

Despite the growing body of literature highlighting the crucial role of ODBs in connecting formal and informal educational domains with employment, several aspects require further investigation: how badges are perceived by learners, institutions, and employers; which badges earners include in their CVs; and how demographic variables influence their adoption (Randall & West, 2022; Roy & Clark, 2019; Stork et al., 2022).

Beginning in the 2018–2019 academic year, INVALSI (Italian National Institute for the Evaluation of the Education and Training System) introduced a certification system for students completing upper secondary education. Competence levels are assigned in Italian literacy, Mathematics, and English (reading and listening), based on national standardised assessments. Italian and Mathematics are evaluated on a five–level scale, while English is assessed according to the Common European Framework of Reference for Languages (CEFR), distinguishing B1 and B2 proficiency levels. To enhance the visibility and usability of these certifications, INVALSI established a digital badging system in 2020 in collaboration with Bestr.it, Italy's leading platform for issuing digital credentials across universities and training institutions.

Data

Starting from the 2020–2021 academic year, students have been able to complete a questionnaire within their personal INVALSI online account prior to requesting ODBs. The questionnaire comprises six closed–ended items covering: prior knowledge of ODBs, perceived usefulness, intention to use, intended usage channels, and post–secondary plans.

Additional information was retrieved from the INVALSI administrative database, including gender, geographical area, type of school attended, and competence level achieved in each test. An original dataset was constructed by merging these two sources using the unique student code as a linkage key. This approach enabled the inclusion of a concise questionnaire (six items only), thereby minimising respondent burden and enhancing response rates (European Commission, 2006, 2011).

Methodology

The analysis comprised both descriptive and inferential components. The descriptive component explored preliminary data pertaining to INVALSI badge earners, with reference to demographic characteristics, attitudes towards ODBs, and intended usage. The inferential component involved the implementation of a binary logistic regression model.

The dichotomous dependent variable represents the propensity to use badges (1=yes, 0=no). The independent variables include competence level attained, gender, geographical area, post–diploma choice, and type of school attended. For each category of the independent variables, an exp(B) coefficient is computed relative to a designated reference category: values greater than 1 indicate higher odds of badge use compared to the reference category; values less than 1 indicate lower odds. Model fit was assessed via the Likelihood Ratio Test, comparing the full model with predictors to the null model.

Results and Discussion

Among students who requested at least one badge, 87.6% (n=60,177) responded in the 2021–2022 academic year. The sample consists of 57.4% female respondents, with a higher concentration in northern regions. The majority of students attend lyceums (60.9%) and express an intention to pursue further studies (80.9%).

As expected, 78.7% of students reported they had never previously encountered ODBs before accessing the INVALSI system, confirming their limited diffusion even among younger generations. The majority of respondents (65.7%) believe badges are useful both for continuing education and for entering the job market. In terms of intended use, English badges attracted the highest interest (approximately 72%), compared to just over half of respondents for Italian literacy and Mathematics badges (57.1% and 56.8%, respectively). This preference is largely driven by the possibility of using INVALSI English badges to obtain university examination exemptions.

Logistic regression results indicate that the probability of using badges increases with the proficiency level achieved; however, all coefficients remain below 1 for levels below the maximum, confirming

that reference categories (Level 5 for Italian and Mathematics, B2 for English) correspond to the highest probabilities.
Contrary to the initial hypothesis, students intending to pursue further studies show a higher propensity to use badges compared to those planning to enter the workforce, especially for English badges (Exp(B)=1.25 for reading and 1.22 for listening). Students from vocational and technical schools demonstrate a higher propensity compared to lyceum students: vocational school students show the highest probability for the Mathematics badge (Exp(B)=1.34), while technical school students show the highest probability for English listening (Exp(B)=1.36). Significant geographical differences emerge, with higher badge use propensity in southern regions. No statistically significant gender differences were observed, although a slight female advantage is present (Exp(B)=1.07).
Regarding intended usage channels, inclusion in the curriculum vitae is the most frequently cited destination (30–35%), followed by presenting English badges for university exemption purposes (approximately 25%).
Conclusions
This study has several limitations, particularly related to the voluntary nature of questionnaire completion, which may have introduced self–selection bias: respondents with lower competency levels appear underrepresented. Nonetheless, the response rate (87.6%) is satisfactory, and the sample size compares favourably with similar studies in the literature (Dyjur & Lindstrom, 2017; Young et al., 2019).
To broaden participation and further promote ODBs as both a tool and a research instrument, the following initiatives are proposed: making questionnaire completion mandatory; enhancing dissemination of INVALSI ODBs across educational and labour market networks, targeting students, teachers, school principals, universities, and businesses; conducting a follow–up survey one year after initial data collection; expanding the number of universities recognising INVALSI English badges as alternatives to traditional proficiency assessments; and continuing the longitudinal linkage of educational administrative data sources, from primary school through university.
The findings largely corroborate existing literature regarding limited ODB awareness. Nevertheless, once acquainted with the tool, users demonstrate willingness to adopt it. This study contributes insights into underexplored aspects — badge perception, badge sharing patterns, and actual usage intentions among students — thus informing strategies for wider dissemination.

Keywords: Competences certification; Digital credentials; INVALSI; Open Digital Badges; High school

SESSION 16. EDUCATIONAL INEQUALITIES, LEARNING OUTCOMES AND OPPORTUNITIES IN CHANGING CONTEXTS

ORGANIZERS: UNIVERSITY OF BERGAMO – UNIVERSITY OF MILAN "LA STATALE"

COORDINATORS: JURGENA MYFTIU – GIULIANA PARENTE

5TH NOVEMBER: 2.00 P.M. – 4.00 P.M. {AULA MAGNA – RESEARCH 10}

Ability–Based Peer Sorting, Ordinal Rank, and Educational Inequality in Academic Achievement

Emanuele Fedeli

Introduction

How do classmates shape students' academic achievement? Research on peer effects has long argued that exposure to high–achieving peers can improve learning by generating positive spillovers, stronger academic norms, and more demanding instructional environments (Lazear 2001; Epple and Romano 1998; Bandura 1977). Yet empirical evidence remains highly inconsistent: some studies find positive peer effects, others report null results, and others even document negative consequences of exposure to stronger classmates. This lack of consensus suggests that classroom peer environments may be more complex than conventional models assume.

This paper argues that one important reason for these mixed findings is that most studies conceptualize peer environments as one–dimensional. Typically, they focus on the average academic ability of classmates and interpret peer effects as learning externalities generated by stronger peers. However, exposure to high–achieving classmates simultaneously changes another feature of students' classroom environment: their relative position within the classroom hierarchy. A student surrounded by high–performing peers may benefit from richer learning opportunities, but at the same time may experience a lower ordinal rank in the classroom, potentially harming motivation, self–concept, and later achievement through social comparison processes (Festinger 1954; Marsh 1987; Davis 1966).

I therefore conceptualize classroom environments as comprising two distinct but interconnected dimensions. The first is a cognitive dimension, captured by the average academic ability of classmates and reflecting the learning opportunities available in the classroom. The second is a positional dimension, captured by students' ordinal rank within the classroom ability distribution and reflecting their relative standing among peers. These two dimensions arise simultaneously from the same classroom composition but may affect achievement through different and potentially opposing mechanisms.

The cognitive dimension is rooted in the classic peer–effects literature in sociology, economics, and educational psychology. High–achieving peers may improve learning by increasing the quality of classroom interactions, raising effort norms, and allowing teachers to adopt a faster or more demanding instructional pace (Lazear 2001; Epple and Romano 1998). From this perspective, exposure to stronger classmates should improve students' later achievement. By contrast, the positional dimension is grounded in social comparison theory and in the literature on the Big–Fish–Little–Pond Effect (Festinger 1954; Davis 1966; Marsh 1987). Students do not evaluate their ability in absolute terms only; they also assess themselves relative to salient reference groups, and the classroom is one of the most important such groups during childhood and adolescence. Holding constant a student's own prior achievement, being assigned to a classroom with stronger peers may lower that student's rank in the local ability hierarchy and thereby reduce academic self–concept, confidence, and persistence.

This framework implies that conventional peer–effect estimates may conflate two countervailing processes: a positive learning effect associated with average peer ability and a negative rank–based effect associated with less favorable relative standing. If these mechanisms operate simultaneously, the net effect of peer ability may appear small, unstable, or even null, even when both underlying channels are substantively important. This argument is consistent with recent work showing that omitting ordinal rank may bias estimates of peer ability by conflating learning externalities with social–comparison effects (Bertoni and Nisticò 2023). The paper therefore asks three related questions: (1) does exposure to higher–achieving classmates improve later achievement once students' classroom rank is taken into account? (2) Does ordinal rank exert an independent effect on achievement net of own ability and peer ability? (3) Do the effects of peer ability and rank vary across students' initial academic ability?

To answer these questions, I use longitudinal administrative data from Italy and follow two national cohorts of students from Grade 5 to Grade 8. A key advantage of the Italian setting is that Grade 5 achievement is observed before students transition into lower secondary school and before they are assigned to new classrooms. This makes it possible to construct peer ability and classroom rank using pre–treatment achievement measures, thereby avoiding simultaneity between peer composition and later outcomes. A second challenge concerns endogenous classroom sorting within schools. To address this issue, I develop a permutation–based within–school sorting test that identifies whether observed classroom allocations are statistically consistent with random assignment. The main analyses are then conducted on the subset of schools whose classroom allocation appears compatible with random sorting.

By jointly modelling average peer ability and classroom rank, this paper contributes to the literature in three ways. First, it provides a multidimensional framework for understanding classroom peer environments, integrating learning spillovers and social comparison in a single model. Second, it offers a design–based strategy to address within–school sorting using administrative data. Third, it shifts attention from the average effect of peers to the way classroom environments may reproduce inequality through the joint distribution of peer ability and relative standing.

Data Used

The empirical analyses rely on administrative longitudinal data from INVALSI, the Italian National Institute for the Evaluation of the Education System, covering two national student cohorts enrolled in Grade 5 in the 2012–2013 and 2013–2014 school years. Students are followed through the end of lower secondary school (Grade 8), allowing achievement to be observed both before and after the transition into new classroom environments.

The timing of the data is central to the research design. In the Italian school system, students are reorganized into new classrooms when they enter lower secondary school. Because Grade 5 test scores are measured before this transition, they can be used to construct pre–treatment measures of classroom environments in lower secondary school. This feature allows the analysis to avoid one of the central problems in the peer–effects literature: the simultaneity between peer characteristics and outcomes when both are measured in the same period (Manski 1993).

The unit of analysis is the student. The analytical sample is restricted to schools with at least two classrooms per cohort, because the empirical strategy relies on within–school comparisons across alternative classroom environments. The outcome of interest is students' standardized Grade 8 achievement, while Grade 5 test scores provide the baseline measure of prior ability.

Classroom environments are operationalized through two complementary measures. The first is average peer ability, defined as the leave–one–out mean of classmates' Grade 5 standardized test scores. This captures the cognitive quality of the classroom environment while avoiding mechanical contamination by the student's own score. The second is ordinal classroom rank, measured as the student's normalized rank in the classroom ability distribution based on Grade 5 scores. Rank ranges from 0 to 1, where higher values indicate a more favorable relative position among classmates.

All models additionally include controls for students' own prior achievement, gender, migration background, cohort, and school fixed effects. To separate the effects of cognitive peer composition from socioeconomic classroom composition, the analysis also controls for the leave–one–out classroom average of the ESCS index.

Methodology

The empirical strategy addresses three major challenges in the peer–effects literature: simultaneity, school–level confounding, and endogenous sorting of students across classrooms.

First, simultaneity is addressed by constructing peer measures using Grade 5 achievement, observed before students enter lower secondary classrooms. Because peer ability and rank are measured using pre–treatment information, they are predetermined with respect to Grade 8 outcomes.

Second, all models include school fixed effects, so identification comes from variation across classrooms within the same school rather than from comparisons across schools. This design absorbs time–invariant school–level differences such as resources, institutional quality, or neighborhood composition.

Third, the key remaining threat is within–school sorting. Schools may deliberately allocate students to classrooms according to prior achievement or socioeconomic background, in which case observed peer environments would not be plausibly exogenous. This concern is central to the peer–effects literature, where non–random assignment of students to peer groups has long complicated causal identification (Manski 1993). To assess this possibility, I implement a permutation–based sorting test. For each school, I compute measures of classroom concentration in prior achievement and socioeconomic status, then repeatedly simulate random classroom allocations while preserving the observed number and size of classrooms. Comparing the observed concentration to the distribution generated under random assignment produces a sorting statistic that captures how unusual the observed allocation is relative to chance. I then use a two–component Gaussian mixture model to classify schools into those whose classroom allocation is broadly consistent with random assignment and those exhibiting stronger–than–random clustering. The main analyses are restricted to the former group.

The baseline empirical model regresses Grade 8 achievement on students' own Grade 5 achievement, average peer ability, classroom rank, and a set of controls. A central specification also includes an interaction between peer ability and rank in order to test whether the consequences of exposure to high–achieving peers depend on students' position in the classroom hierarchy. Comparing models with and without rank provides a direct way to assess whether conventional peer–effect estimates mask the coexistence of learning and social–comparison mechanisms (Bertoni and Nisticò 2023).

Results and Discussion

The paper advances three main findings.

First, the results suggest that the effect of peer ability cannot be interpreted as a single, one–dimensional treatment. Once ordinal classroom rank is introduced into the model, the estimated effect of average peer ability changes, indicating that conventional peer–effect estimates combine at least two distinct processes. This pattern is consistent with the idea that high–achieving peers simultaneously provide richer learning environments and reduce some students' relative standing within the classroom hierarchy.

Second, classroom rank appears to matter independently of average peer ability and students' own prior achievement. This result supports a social–comparison interpretation of classroom environments: students are affected not only by who their classmates are, but also by where they stand relative to them. In other words, two students with identical prior achievement may experience different academic trajectories if one occupies a relatively favorable position in the classroom while the other is placed near the bottom of a stronger class. This finding is in line with the literature on social comparison and the Big–Fish–Little–Pond Effect, which emphasizes how relative standing in a salient reference group shapes academic self–concept and performance (Festinger 1954; Davis 1966; Marsh 1987).

Third, the relationship between peer ability and later achievement appears to be heterogeneous rather than uniform. The inclusion of an interaction between average peer ability and rank is motivated by the hypothesis that exposure to high–achieving peers may be more beneficial for students who retain relatively favorable positions within the classroom hierarchy, while being less beneficial—or even harmful—for those who experience a strong loss in relative standing. This possibility offers a compelling explanation for why the literature has often produced mixed or weak average peer effects: average estimates may conceal offsetting mechanisms that operate differently across students.

More broadly, the results help reconcile two strands of research that have often been treated separately: the literature on positive peer spillovers and the literature on rank–based social comparison. The first emphasizes how stronger peers improve the quality of the learning environment (Lazear 2001; Epple and Romano 1998); the second shows how local rank shapes academic self–concept and performance (Festinger 1954; Marsh 1987). This paper shows that both mechanisms can operate simultaneously and that their interaction is crucial for understanding the overall consequences of classroom peer composition.

Conclusions

This paper develops a multidimensional framework for classroom peer environments by distinguishing between a cognitive dimension, captured by average peer ability, and a positional dimension, captured by students' ordinal rank within the classroom. Using longitudinal administrative data from Italy and a design that combines pre–treatment peer measures, school fixed effects, and a permutation–based test for within–school sorting, the study shows that conventional peer–effect estimates may obscure the coexistence of positive learning externalities and negative social–comparison mechanisms. The findings suggest that classroom environments shape achievement not only through exposure to stronger or weaker peers, but also through students' relative standing within those environments, with important implications for both research and policy: models of peer effects should account explicitly for rank, and policies of tracking, mixing, or classroom assignment should be evaluated not only in terms of average peer composition but also in terms of how they structure students' relative positions and, in turn, educational inequality.

Keywords: Peer effects; Classroom rank; Social comparison; Educational inequality

Patterns and Mechanisms of Educational Inequality across Europe: New Evidence and Trends

Ilaria Lievore

Introduction

Educational inequalities remain one of the most persistent and policy–relevant challenges facing European education systems. A large body of comparative research has documented substantial achievement gaps by socioeconomic background, gender, and migration background, especially during adolescence. Much of this evidence relies on large–scale cross–national assessments such as PISA, TIMSS, and PIRLS, which have played a crucial role in identifying cross–country differences in student achievement and in highlighting the relationship between social background and educational outcomes. However, the predominantly cross–sectional nature of these surveys limits our ability to understand educational inequalities as dynamic processes that unfold over time. From a comparative perspective, several fundamental questions therefore remain only partially answered. When do achievement gaps first emerge? Are inequalities already present before formal schooling, or do they develop during primary and secondary education? How do inequalities evolve across different stages

of compulsory schooling? To what extent do national education systems shape the timing, persistence, or widening of these gaps?

These questions are central not only to academic debates on social stratification and educational opportunity, but also to the design of effective and timely policy interventions. If inequalities emerge early and remain stable, policy responses should focus on early childhood and primary education. If they widen during specific educational transitions, interventions should target those institutional turning points. Addressing these issues requires longitudinal data that follow students across multiple years, grades, and educational transitions. Although the availability of longitudinal education data has increased in recent years, such data remain unevenly distributed, fragmented, and often underused across Europe. This is particularly true for several Southern European countries, which remain underrepresented in comparative longitudinal research on educational inequality despite having distinctive institutional structures and policy contexts.

This work addresses this gap by providing descriptive comparative evidence on the evolution of educational inequalities across seven European education systems. The study draws on national longitudinal or longitudinally linkable data, harmonised ex post within the LINEup Horizon Project. Its main objective is to compare how achievement gaps by gender, migration background, parental education, and parental occupation evolve across different stages of schooling.

Data Used

The study uses national longitudinal education datasets from seven European countries or education systems: Germany, Greece, Catalonia, France, Italy, Portugal, and the United Kingdom. These datasets differ in design, coverage, available outcomes, and institutional context, but they all allow students to be observed across more than one grade or educational stage.

For Germany, the study uses data from the National Educational Panel Study, specifically two cohorts. The Kindergarten Starting Cohort, launched in 2010, follows children from early childhood education into adulthood. The Grade 5 Starting Cohort, also launched in 2010, follows students from early adolescence into adulthood. These cohorts include approximately 6,000 and 5,000 students respectively and provide information on both competence tests and teacher grades.

For Greece, the study uses data from MySchool, the unified information system developed by the Ministry of Education, Religious Affairs and Sports to support the digital management of school units and educational administration. The selected cohort follows students from Grade 7 to Grade 12 between the 2019/2020 and 2024/2025 academic years. The cohort includes approximately 100,000 students at the beginning of the observation period and around 66,000 students in the final year.

For Catalonia, the study uses the RALC student registry, maintained by the Catalan Department of Education. RALC covers the full student population of the Autonomous Community of Catalonia. The selected cohort consists of students who first entered Grade 6, the final year of primary education, in the 2017/2018 academic year. These students are followed for five years, until 2021/2022, when they are expected to have completed Grade 10. The cohort includes approximately 80,000 students.

For France, the study uses longitudinal panels produced by the Direction de l'évaluation, de la prospective et de la performance under the French Ministry of Education. Two cohorts are included. The 2007 panel follows students from entry into Grade 6 through secondary education and beyond. The 2011 panel follows students who entered Grade 1 in 2011 and is still ongoing. The selected data include approximately 35,000 students at baseline in the 2007 panel and 15,200 students in the 2011 panel. Educational outcomes include standardised tests in literacy and numeracy.

For Italy, the study uses data from INVALSI, the National Institute for the Evaluation of the Education System. INVALSI data are collected through periodic standardised assessments of students' knowledge and skills. Although the data are collected as repeated cross–sections, students can be linked across grades and academic years through a unique identifier. Three cohorts are selected: students followed from Grade 5 in 2013/2014 to Grade 13 in 2021/2022; students followed from

Grade 2 in 2015/2016 to Grade 10 in 2023/2024; and students followed from Grade 2 in 2014/2015 to Grade 10 in 2022/2023.

For Portugal, the study uses data from the student register managed by DGEEC. The dataset combines information from three administrative sources: MISI, Escola 360, and CIF. Together, these sources allow the construction of longitudinal student trajectories across school years. The selected cohort includes approximately 105,000 students enrolled in Grade 5 in the 2014/2015 academic year and follows them until 2021/2022, when they reached Grade 12.

For the United Kingdom, the study uses the Millennium Cohort Study, a longitudinal birth cohort study following approximately 15,000 children born between 2000 and 2002. The study provides longitudinal information on children's educational development and family background.

Methodology

The study follows an ex–post, retrospective harmonisation strategy. Because the datasets were not originally designed as part of a common international survey, harmonisation was conducted after data collection. The harmonisation process selected a common set of variables and recoded them according to a minimum harmonisation logic. This approach preserves the largest possible number of countries and cohorts while ensuring that the key measures are sufficiently comparable for descriptive cross–country analysis.

The harmonisation process was conducted within the LINEup Horizon Project and involved several partner organizations. Since not all microdata are directly accessible to the authors of this work, the final harmonised dataset consists of aggregated estimates rather than pooled individual–level data. Each partner organization produced estimates using a common analytical protocol and shared the resulting aggregate outputs with the authors.

The harmonised independent variables are gender, migration background, parental education, and parental occupation. Parental education is measured separately for mothers and fathers, using the highest level of education available in each dataset. Parental occupation is also measured separately for mothers and fathers, using harmonised occupational status categories. The harmonised dependent variable is student performance. Depending on data availability, this is measured through standardised test scores, teacher grades, or both. Because scoring systems differ substantially across countries, datasets, subjects, and grades, the analysis relies on within–dataset standardised measures of performance. Scores are standardised within cohort, academic grade, academic year, and subject. This strategy allows the analysis to compare relative student performance within each education system, rather than absolute achievement levels across countries.

The main estimates are group differences in standardised performance at each observed grade, calculated separately by country, dataset, cohort, subject, and wave. For each dimension of inequality, partner organizations estimate mean differences in performance between relevant student groups. These estimates are produced using population weights where available. When original student weights are not available, partner organizations are asked to construct appropriate weights to account for unbalanced samples, attrition, or missing information. Standard errors are clustered at the student level to account for repeated observations of the same individuals over time. For each estimate, partner organizations report adjusted margins, standard errors, confidence intervals, group sizes, country, grade, wave, cohort, and subject. They also provide a technical note describing cohort selection, sample restrictions, treatment of missing data, outcome construction, and weighting strategy. These technical notes are used to assess the comparability of the estimates and to document country–specific limitations.

Once the country–level aggregate estimates are received, they are combined into a single harmonised dataset. This dataset allows the analysis to describe and compare trajectories of educational inequality across countries, cohorts, grades, and subjects. The analysis remains descriptive and does not aim to identify causal effects. Instead, it provides a harmonised comparative overview of when achievement gaps appear, whether they widen or narrow over time, and how these patterns differ across education systems

Results and Discussion

As the empirical analysis is currently ongoing, this section presents the planned analytical strategy and the type of evidence that the study will provide. The expected results will describe the evolution of achievement gaps by gender, migration background, parental education, and parental occupation across the available grades and cohorts in each country.

First, the analysis will examine whether achievement gaps are already visible at the earliest observed grade in each dataset. This will make it possible to assess whether inequalities appear before or at the beginning of formal schooling, during primary education, or later in students' educational trajectories. Second, the study will analyze whether inequalities remain stable, increase, or decrease as students progress through school (within–country trends). For each country and cohort, group differences in standardised performance will be plotted by grade and subject. This will allow to identify periods or transition points in which inequalities widen or narrow in relative terms, including key educational transitions such as the move from primary to lower secondary education and from lower to upper secondary education. Third, we will assess whether similar forms of inequality follow comparable patterns across education systems, or whether their timing and development differ substantially by national context. This comparative perspective is especially relevant for countries that are often absent from longitudinal studies of educational inequality, including Greece, Portugal, and Catalonia. Fourth, the study will examine whether different dimensions of inequality follow similar or distinct trajectories.

The discussion will interpret these descriptive patterns in relation to institutional features of education systems, such as the timing of school transitions, the structure of tracking, and the availability of compensatory policies, identifying cross–country patterns that can inform future explanatory research and guide the development of more targeted policy interventions.

Conclusions

This study contributes to comparative research on educational inequality by using harmonised longitudinal evidence from seven European countries. It responds to a key limitation of existing comparative research, which mainly relies on cross–sectional assessments overlooking how inequalities develop over time across different countries.

The study makes three main contributions. First, it shifts the focus from static achievement gaps to trajectories of inequality across schooling. Second, it brings together national longitudinal and longitudinally linkable datasets that are rarely analyzed within a common comparative framework. Third, it gives visibility to Southern European education systems, which remain underrepresented in the longitudinal literature despite their relevance for understanding institutional variation in Europe.

The analysis will provide descriptive evidence on when achievement gaps emerge, how they evolve, and whether their trajectories differ across countries and dimensions of inequality. By doing so, it will help identify the educational stages at which inequalities are most pronounced and the moments at which policy interventions may be most effective.

At the same time, the study has important limitations. The harmonisation process relies on minimum common definitions, which necessarily reduces the complexity of some national measures. The use of aggregate estimates rather than pooled microdata also limits the range of possible analyses. Moreover, differences in outcome measures, cohort design, grade coverage, and data quality require careful interpretation of cross–country comparisons. While the harmonisation of different longitudinal datasets, which are built for different purposes, has the major drawback of loss of information, it allows descriptive analytical potential for advancing comparative longitudinal research on educational inequality. It also highlights the need for greater accessibility, documentation, and comparability of national education data across Europe.

Keywords: Educational inequality; Longitudinal data; Comparative education; Achievement gaps

Immigrant–Native Disparities in Bullying and School Belonging: A Multilevel Socio–Ecological Analysis Across Europe

Ioulia Televantou – Herb Marsh – Theresa Dicke

Introduction

There is a longstanding view that schools function as a "microcosm of society," reflecting broader social norms, inequalities, and attitudes in students' everyday interactions (McCoy, 2026). This perspective raises a critical question: to what extent do societal–level factors, such as public attitudes toward immigrants and national integration policy frameworks, align with immigrant–native disparities in students' social experiences at school, particularly in bullying victimization and school belonging?

The present study draws on cross–national three–level data from two complementary sources, the OECD Programme for International Student Assessment (PISA 2022) and the European Social Survey (ESS Round 10) and adopts a social–ecological perspective to examine how immigrant–native disparities in bullying victimization and school belonging vary across individual, school, and national contexts. Bullying victimization and school belonging are examined jointly as complementary dimensions of adolescents' social experiences within school contexts, capturing both exposure to negative peer interactions and the extent to which students experience inclusion and connectedness within the school community. The analyses provide a multilevel mapping of how these disparities are patterned across schools and countries. By situating immigrant status within indicators of school context and national integration climates, the study delineates the broader environments within which adolescents' peer experiences unfold.

At the individual level (Level 1; microsystem), we compare immigrant adolescents with native–born peers and examine whether disparities in bullying victimization and school belonging persist after accounting for demographic and socioeconomic characteristics. At the school level (Level 2; mesosystem), we assess whether contextual features, namely the school immigrant composition and the school principal's diversity and multicultural views (M–SC), moderate immigrant–native differences in students' social experiences. At the country level (Level 3; macrosystem), we investigate whether national sociocultural environments, specifically public attitudes toward immigrants and integration policy frameworks, correspond to cross–national variation in these disparities. Through this multilevel design, the study evaluates whether school and societal contexts are associated with the magnitude of immigrant–native gaps in bullying victimization and school belonging.

The study extends prior research in several ways. First, it revisits the hypotheses examined by Stevens et al. (2020) using more recent data, refined operationalizations of bullying victimization and societal attitudes, and a broader socio–ecological framework that incorporates school belonging alongside bullying victimization while simultaneously considering both school– and country–level contexts. Second, it complements the work of Basarkod et al. (2024), who demonstrated the importance of intercultural classroom practices and teachers' openness for student well–being. Whereas that, and more recent research (see McCoy et al., 2026) highlights microsystem and mesosystem processes within schools, the present study focuses on how such processes intersect with students' immigration background and with contextual factors operating at the macrosystem level.

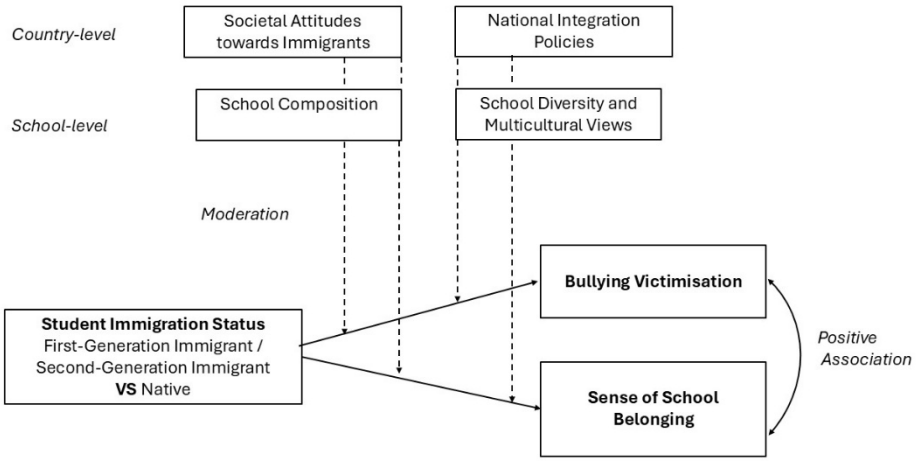


Figure 1. Conceptual model of immigrant youths' bullying victimization and school belonging.
Bullying victimization and sense of school belonging are conceptualized as correlated outcomes. Student immigration status is theorized as the central individual–level predictor, whereas school–level (L2) and country–level (L3) contextual characteristics are conceptualized as moderators of the association between immigration status and both outcomes. Student background characteristics (socio–economic status, gender, age, number of siblings, language spoken at home) are included as controls and are not the primary theoretical focus

Research Questions and Research Hypothesis

Research Question 1. To what extent are immigrant–native differences in bullying victimization and school belonging observable after adjusting for individual–level characteristics (e.g., socioeconomic status, gender, age, language spoken at home, and number of siblings)?

*Research Question 2***.** To what extent do immigrant–native differences in bullying victimization and school belonging vary across schools and countries?

Research Question 3. To what extent do immigrant–native differences in bullying victimization and school belonging correspond to variation in school–level composition and country–level contexts, namely societal attitudes toward immigrants?

Research Question 4. To what extent do immigrant–native differences in bullying victimization and school belonging correspond to variation in school climate (including diversity and supportive climate) and educational policy contexts (MIPEX–Edu indicators)?

Methodology

Data and Sample Size

We used representative samples of fifteen–year–old students from thirty European countries (see vertical axis labels, Figure 2). The data are drawn from the PISA 2022 assessment frameworks; we include only European countries that also participated in ESS10 in our sample. The data are hierarchical, with students nested within schools, and schools nested within countries; this three–level structure was accounted for in subsequent statistical analyses. The analyses were based on 217,913 students from 7,829 schools across 30 European countries. The average school size was approximately 28 students, and the average number of students per country was about 7,264.

Measures

Individual–level measures were drawn from the PISA 2022 student questionnaire. These variables are available as standardized composite IRT scores (OECD, 2024, p. 398). It should be noted that the measurement invariance of the bullying victimization and school belonging scales across immigrant and non–immigrant students in PISA 2022 has already been examined by Casas and Sireci (2025), who reported acceptable model fit in support of measurement invariance for both constructs.

Following Basarkod et al. (2022) and OECD (2023), first–generation immigrant students were defined as students who were born abroad and whose parents were also both born abroad. Second–generation immigrant students were defined as students born in the host country whose parents were both born abroad. Native students were defined as students with at least one parent born in the host country, regardless of whether the students themselves were born in the host country or abroad. At the school level, immigrant composition was operationalized as the proportion of first–generation immigrant students enrolled in each school. School diversity and multicultural views were measured using principals' responses to a set of items capturing the extent to which school staff promoted a diversity–oriented culture and climate during the previous academic year (OECD, 2024).

At the societal level, citizens' attitudes toward immigrants were derived from the European Social Survey (ESS Round 10) and aggregated to the country level. Two country–level constructs were created. The first, "Allowing Immigrants in the Country," measured public support for immigration and included three items assessing whether immigrants from different backgrounds should be allowed to live in the country. The second, "Quality of Life and Immigrants," captured broader evaluations of the societal consequences of immigration and included three items assessing whether immigration is good or bad for the country's economy, whether immigrants undermine or enrich cultural life, and whether immigrants make the country a better or worse place to live. The main analyses used the "Quality of Life and Immigrants" construct as the primary indicator of societal attitudes toward immigrants, while the "Allowing Immigrants in the Country" construct was used in supplementary analyses to assess the robustness of the findings. The internal consistency of the ESS attitudinal measures was assessed using Cronbach's alpha and omega reliability, while construct validity was examined through multigroup confirmatory factor analysis (MGCFA; Jöreskog, 1971).

Finally, the Education–Specific Migrant Integration Policy Index (MIPEX–Edu) was used as a country–level indicator of education policy frameworks. For the purposes of the present study, we used the MIPEX–Edu edition published in 2020, as this was the closest available edition to the PISA 2022 and ESS Round 10 data collections (Solano & Huddleston, 2020). Higher MIPEX–Edu scores indicate more extensive policy provision in relation to migrant education. However, the index does not capture how such policies are implemented or experienced in everyday school contexts.

Statistical Analysis

Data preparation was conducted using IBM SPSS Statistics, while all statistical analyses were performed in Mplus version 8.6. Missing data were addressed using Full Information Maximum Likelihood (FIML), which is the default estimation method in Mplus. All analyses incorporated the PISA final student weights to account for the complex sampling design. This approach is consistent with prior PISA–based multilevel research examining student–, school–, and country–level associations. (Karakolides et al., 2022). A multivariate three–level model was estimated to examine immigration background and student characteristics in relation to bullying victimization and school belonging as correlated outcomes.

Findings

Descriptive analyses indicated substantial cross–country variation in both bullying victimization and school belonging Across countries, higher average levels of bullying tended to coincide with lower levels of belonging, and within–country variability was consistently greater among immigrant students. First–generation immigrant students reported higher levels of victimization and lower levels of belonging than native peers across most countries.

Bullying victimization and school belonging were strongly and inversely related across all levels of analysis. At the individual level, students who reported higher levels of bullying also reported lower belonging ($r = -.252$, $p < .001$). This pattern was consistent at the school ($r = -.451$, $p < .001$) and country levels ($r = -.438$, $p < .001$). Intraclass correlation coefficients indicated that a modest proportion of variance was attributable to contextual levels. For bullying victimization, 3.3% of variance was at the school level and 2% at the country level. For school belonging, corresponding values were 3.6% and 2%, respectively.

Immigrant–Native Differences

Models including immigration background indicated that first–generation immigrant status was associated with higher bullying victimization (b = .031, SE = .022, p < .001; β = .142) and lower school belonging (b = −.059, SE = .028, p < .001; β = −.264). Second–generation immigrant status was also associated with bullying victimization (b = .015, SE = .023, p = .017; β = .055) and school belonging (b = −.028, SE = .033, p = .002; β = −.104). After adjustment for student–level characteristics, first–generation immigrant status remained associated with bullying victimization (b = .052, SE = .021, p = .015) and school belonging (b = −.158, SE = .027, p < .001). Second–generation immigrant status was not statistically significant for either bullying victimization (p = .131) or school belonging (p = .061) after adjustment, including language spoken at home.

Variation Across Schools and Countries

Random–slope models indicated that the association between first–generation immigrant status and bullying victimization varied significantly across schools (Var = 0.055, SE = 0.016, p = .001), whereas no such variation was observed for school belonging ($p = .992$). At the country level, the immigrant–native gap varied significantly for both bullying victimization (Var = 0.011, SE = 0.005, p = .029) and school belonging (Var = 0.011, SE = 0.003, p = .001). These findings indicate that immigrant–native disparities differ systematically across both schools and countries.

School and Societal Contexts

For bullying victimization, multilevel models with random slopes for first–generation immigrant status were estimated at both the school and country levels, while school belonging was retained as a parallel outcome to account for correlations between the two constructs. At the school level, a higher proportion of first–generation immigrant students was associated with a smaller immigrant–native gap in bullying victimization ($\beta = -.336, SE = .148, p = .024$). At the country level, more positive public attitudes toward immigrants, measured through evaluations of immigrants' contribution to quality of life, were also associated with smaller disparities in bullying victimization (β = −.089, SE = .019, p < .001). For school belonging, school immigrant composition was not associated with variation in immigrant–native disparities across schools (β = .114, SE = .103, p = .272), and the corresponding slope variance was not statistically significant. This suggests that although school composition relates to average levels of belonging, it does not explain differences in belonging between immigrant and native students across schools. In contrast, public attitudes toward immigrants were significantly associated with variation in immigrant–native disparities in school belonging (β = −.054, SE = .020, p = .007). Specifically, in countries where immigrants are viewed more positively in terms of their contribution to quality of life, disparities in school belonging were smaller. Significant residual slope variance further indicated meaningful cross–national heterogeneity.

School Climate and Policy Contexts

In the second set of analyses. we examined cross–level moderation of the effect of first–generation immigrant status by diversity–supportive and multicultural school climate (M–SC) at the school level and by inclusive education–related integration policies at the country level (MIPEX–Edu; see moderation paths in the right in Figure 1), controlling for school composition and societal attitudes towards immigrants. At the school level, diversity–supportive and multicultural school climate (M–SC) was not associated with between–school variation in immigrant–native disparities in bullying victimization (β = −0.009, SE = 0.014, p = .530) or school belonging (β = 0.003, SE = 0.010, p = .774). At the country level, MIPEX–Edu showed no significant main effects for either bullying victimization (p = .962) or school belonging (p = .143). However, MIPEX–Edu was marginally associated with variation in immigrant–native disparities in bullying victimization (β = −0.002, SE = 0.001, p = .049). Specifically, in countries with higher MIPEX–Edu, disparities in bullying victimization between first–generation immigrant and native students were smaller. MIPEX–Edu was also significantly associated with variation in immigrant–native disparities in school belonging (β = −0.002, SE = 0.001, p = .044). The fixed effect of first–generation immigrant status on school belonging was negative and

statistically significant (β = −0.120, SE = 0.046, p = .008). Thus, the negative association between MIPEX–Edu and the random slope indicates that the immigrant–native gap in school belonging becomes larger in magnitude in countries with more inclusive integration policies.

Discussion

The present study was grounded in a socio–ecological framework (Bronfenbrenner, 1979, 1994; Bronfenbrenner & Morris, 2006), which provides a comprehensive basis for examining contextual variation in students' experiences of bullying victimization and sense of school belonging across individual, school, and national levels. Across thirty European countries, the findings indicated that, on average, first–generation immigrant students experienced higher levels of bullying victimization and a lower sense of school belonging than native students, even after controlling for socioeconomic status and language spoken at home. In contrast, second–generation immigrant students did not differ significantly from native students once language was taken into account, suggesting that linguistic integration may play an important role in students' social adjustment. Socioeconomic status was unrelated to bullying victimization but positively associated with school belonging. At the macrosystem level, which was the main focus of the present study, we distinguished between societal attitudes toward immigrants and education–related integration policy frameworks. While the former capture shared societal orientations, the latter reflect formal policy structures shaping the institutional context of immigrant inclusion. Societal attitudes were consistently associated with cross–national variation in immigrant–native disparities, particularly in bullying victimization. Across both operationalisations of public attitudes, "Allowing Immigrants in the Country" and "Quality of Life and Immigrants," more positive societal views were associated with smaller disparities in bullying victimization between first–generation immigrant and native students. The convergence of findings across these two measures underscores the robustness of this pattern and suggests that broader societal climates surrounding immigration, although distal, are closely linked to immigrant students' peer experiences at school. For school belonging, societal attitudes were also associated with variation in the immigrant–native gap, although these associations were smaller in magnitude and less consistent. Education–related integration policy frameworks were also associated with cross–national variation in immigrant–native disparities, but in different directions depending on the outcome. More comprehensive policy frameworks were associated with smaller disparities in bullying victimization, whereas for school belonging they were associated with a larger negative gap. However, both effects were very small in magnitude. One possible interpretation is that formal policy frameworks capture the institutional design of integration systems but may not fully reflect how such policies are implemented or experienced in everyday school contexts. As a result, the link between policy frameworks and students' lived experiences may be indirect and contingent on how policies are translated into practice at the school level, including variation in school resources, leadership, and teacher practices.

Keywords: Bullying Victimization; School Belonging; Immigration Background; Social–Ecological Framework

The Role of School Type in Shaping Literacy Skills: Evidence from the UK Millennium Cohort Study

Giulietta Zanga – Ekaterina Kochergina

Introduction

Worldwide educational inequalities remain one of the central challenges facing education systems. Despite substantial public investment in education, large differences in academic achievement persist. Existing research highlighted the importance of educational environments in shaping

children's academic trajectories. The effect of school type has attracted particular attention in the UK because of the peculiarities of its education system. Both within the state sector and in the private one, there is significant diversity — from local authority–maintained schools to academies, free schools, faith schools, and special schools (Department for Education, 2026), but also prep schools, independent schools and boarding schools. These schools differ in their governance, admissions policies, student composition, and available resources, potentially leading to differences in educational outcomes (Department for Education, 2026).

Jerrim and Sims (2019) using the UK Millennium Cohort Study (MCS) examine selective schooling and find that school type may be associated with different academic advantages among students attending them. Similarly, Beshir et al. (2024), using linked MCS and National Pupil Database data, demonstrate that although educational inequalities are strongly associated with differences in family resources, school environment can mitigate this association, raising important questions about the role of different school sectors in promoting or limiting educational opportunity. Finally, recent evidence suggests that educational inequalities widen during secondary school. Jerrim and Palma Carvajal (2025) show that high–ability pupils from disadvantaged backgrounds begin to fall behind similarly able peers from more advantaged families during adolescence, indicating that the secondary school years represent a critical period for the development of educational inequality.

Literacy skills constitute an ideal outcome through which to examine these questions. Reading and language proficiency are foundational competencies that influence later educational attainment, labour market success, and social mobility (Meehan et al., 2023; OECD, 2012). Moreover, literacy develops continuously throughout childhood and adolescence, making it possible to investigate how educational environments shape skill development over time (Borgonovi & Pokropek, 2021).

This study is a part of the LINEup project (LINEup project, 2024), which aims to strengthen the evidence base on educational inequalities by mapping and analysing longitudinal data on students' learning outcomes across primary and secondary education in Europe. In line with LINEup's research objectives this study examines whether attendance at different types of schools is associated with differences in students' literacy development. Literacy is used as the main outcome due to its fundamental role in educational attainment, employability, and social inclusion.

The longitudinal design of the UK Millennium Cohort Study (MCS) enables the analysis of literacy development across multiple stages of childhood and adolescence. By leveraging rich longitudinal information on family background, early cognitive skills, and demographic characteristics, the study contributes to the broader goal of disentangling how educational environments interact with prior disadvantage. The analysis therefore provides evidence on whether differences in literacy development across school types reflect the role of school systems themselves, or pre–existing differences in student composition, thereby contributing to policy–relevant insights on equity and school effectiveness across education systems.

Dataset

The Millennium Cohort Study is a longitudinal birth cohort study conducted by the University College of London. Its main objective is to follow the lives of children born between 2000 and 2002 in the United Kingdom. The study follows a pure longitudinal design and initially included approximately 19,000 participants. It covers the entire United Kingdom, including England, Wales, Scotland, and Northern Ireland. Data collection began in 2000 and is still ongoing. Educational outcomes are measured through standardized tests in literacy and numeracy, as well as school performance indicators.

The primary outcome variable in this study will be students' literacy ability, measured using the standardized tests assessing the linguistic domain administered during the survey waves. At age 7, children were tested on their ability to read a series of words presented on a card, at age 11 they were given three stimulus words and asked to name the class to which all the examples belong, while at age 14 they had to pick the word with the same meaning as the one showed to them. The opportunity to track students' cognitive abilities at three different points in time, when they are both

in primary and in secondary school, allows us to verify whether children's characteristics or school type have a different impact on students' assessments at different ages.
The principal explanatory variable is school type attended, public or private school. In the macro category of public schools, we include schools families have not to pay a fee for their children to go to, while private schools include institutions which require fees. The analysis will examine whether literacy outcomes differ across these public and private schools, after accounting for observable differences among students.
A comprehensive set of control variables will be included to reduce potential confounding. These will include child characteristics (such as sex, NUTS region of residence and age), and family socioeconomic status (including household income, parental occupation, and family structure).
For this study, we use a sub–dataset of the whole MCS dataset, which is the result of a process aimed at harmonizing relevant variables that can explain educational inequalities in different European datasets, from the Horizon project LineUp. This dataset includes variables describing students and families' socio–economic characteristics and some school characteristics too, but does not include the thousands of variables collected by MCS, which are not all available to researchers.

Methodology

Multivariate linear regressions have been performed in order to verify a possible correlation between school type and literacy results, net of students' characteristics. Being MCS a longitudinal dataset, our analysis distinguishes results according to the three waves at our disposal (age 7, age 11 and age 14), at the same time looking for possible patterns over time. At the time of the conference also new estimates based on matching techniques will be provided.

Results and discussion

Even if students attending private schools are a small part of our sample, around the 6% of each wave, results clearly and significatively state that children attending private schools perform better than children attending public school, even net of relevant characteristics often associated with dis/advantages in school outcomes.
With no significant differences between waves, children in public school perform around 0.32 standard deviations lower than their peers in private schools on the linguistic outcome ($p < 0.001$), also after controlling for significant characteristics such as sex, parental occupation, income and region of residence.
This means that private schools may have a positive effect on their students' outcome that goes beyond the one of their socio–cultural environment. Additional analyses will be developed to make this causal claim more robust (beyond matching, i.e. controlling for additional features available in the dataset).
Parental occupation is another variable which shows a correlation with children's linguistic outcomes, meaning that students whose parents are unemployed or manual workers perform worse than their peers whose parents are high or low professionals. At the same time, the effect of parental occupation does not reduce the significance of attending a private school, meaning that for each category of parental occupation children attending private schools perform better than children in public school whose parents have similar kinds of jobs. This is true for four categories of parental education (Manual workers, Manual self–employed workers, Low professional and High professionals) but not for children living in families with both unemployed parents (for whom the sample is very small, as predictable).
We could speculate that better school resources, in terms of economic, symbolic and social capital, give children attending private schools advantages that go beyond their own family economic resources, but we cannot disentangle these influences with our data, which focus way more on students' and their families' characteristics than on the schools' ones.

Conclusions

The results coming from our dataset from the Millennium Cohort study are not consistent with previous research that observed that much of the performance difference between students

attending public and private schools reflects the characteristics of the students rather than the ones of the school attended.
Private schools have indeed a positive effect on students' language outcomes, and this is relevant as mainly students coming from wealthy families have the resources and opportunities to attend them, so that economic inequalities in the UK are reproduced also by the types of school institutions.

Keywords: Educational inequality; School type; Literacy development; UK Millennium Cohort Study (MCS)

Mapping Educational Transformation: A Longitudinal Institutional Evaluation of Higher Education Teaching Innovation Projects (2017–2025)

Arantza Arruti

Introduction

Rapid societal shifts force higher education institutions to implement transformative, high–quality educational models and to face some challenges (Arruti and Paños–Castro, 2025). Rather than rely on sporadic individual efforts, forward–thinking universities design internal competitive calls for Teaching Innovation Projects (TIP) to systematize pedagogical modernization. However, evaluating these multi–year initiatives longitudinally is vital to ensure institutional accountability, measure resource efficiency, and prevent the short–lived "innovation bubble".
This study is based on the need to reflect on teaching practice, innovate, and conduct research on it—as one of the tasks that university faculty must carry out in order to contribute to the improvement of their own teaching practice and students' learning experience, and to achieve a transformative education (Castillo Téllez and Pedraz–Nájar, 2023). It also stems from the need to innovate in order to contribute to the quality of teaching, the development and transformation of institutions of higher education, and the improvement of teaching practice. The study is theoretically grounded in the Scholarship of Teaching and Learning (SoTL) framework (Albergaria Almedia, 2010; Paricio, 2020). SoTL positions university instruction as a high–level academic endeavor subject to rigorous empirical inquiry, systematic reflection, and evidence–based refinement. By treating the classroom as an active site of educational research, faculty members shift from being passive presenters to reflexive pedagogical practitioners.
Furthermore, true pedagogical innovation functions as an intentional, deliberative process of critical reflection that challenges traditional lecture paradigms (Mancilla, 2023). According to Feixas and Martínez–Usarralde (2022), a project qualifies as a driver of transformative education when it incorporates interdisciplinarity, research–oriented designs, collaborative learning networks, and alternative learning spaces. To sustain this transformation, contemporary literature highlights that institutional calls must align with systemic quality assurance systems, evolving student profiles, and urgent global challenges. These challenges include the integration of United Nations Sustainable Development Goals (SDGs) and the rapid, ethical deployment of Generative Artificial Intelligence (GenAI) across academic curriculums.
This research conducts a longitudinal case study at the University of Deusto (Spain), an institution recognized for its competence–based educational model (the Deusto Learning Model or MAUD). The study evaluates how competitive funding, institutional support structures, and strategic university plans (2018–2022 and 2023–2026) interact to reshape university teaching over an eight–year period.

Data Used

The data corpus for this evaluation originates from the university's internal quality assurance and pedagogical support registry, spanning eight consecutive institutional funding cycles (TIP calls IX through XVI, from the 2017/2018 to the 2024/2025 academic years).

The empirical dataset comprises three distinct document categories, allowing for multi–source triangulation:

1. **Competitive Institutional Proposals (n = 146):** The complete set of officially approved project applications, representing an initial historical population of 195 submissions across six distinct university faculties and interdisciplinary services.
2. **Final Project Reports (n = 114):** The official final documentation compiled and submitted by leading faculty members at the conclusion of their funded cycles.
3. **Institutional Annual Audit Reports (n = 8):** The comprehensive evaluation evaluations authored by the university's internal Teaching Innovation Unit (*Unidad de Innovación Docente*).

Methodology

This study utilizes a **mixed–methods research design**, integrating descriptive non–experimental quantitative analysis with qualitative content analysis. This combination provides a holistic, data–driven perspective on institutional change.

The quantitative phase applies descriptive statistics to map structural metrics across the eight–year timeline. We analyzed project approval rates (74.8% global acceptance across 195 submissions), budget efficiency, and the distribution of innovation activities across different academic disciplines. Interfacultative/interdisciplinary proposals formed the largest cohort (38.31%), followed by the Faculty of Education and Sport (24.65%) and the Faculty of Social Sciences and Humanities (15%).

The qualitative phase uses structured, matrix–based thematic content analysis to scrutinize the project proposals, final reports, and annual audits. Codes are classified into four evaluative dimensions:

- *Typology of pedagogical intervention* (e.g., active methodologies, assessment strategies, digital virtualization).
- *Degree of objective achievement*.
- *Sustainability* (the project's survival rate beyond the funded institutional cycle).
- *Pathway of institutional transferability* (classified as either localized departmental adaptation or trans–university, systemic impact).

Results and Discussion

The structural data demonstrates a highly selective and selective institutional framework. The strict review mechanisms used by internal evaluation committees ensured that resources were prioritized for sustainable, methodologically sound project designs.

The qualitative evaluation reveals three major trends in higher education innovation:

1. **The Evolution toward SoTL:** The early funding cycles focused heavily on isolated, small–scale classroom experiments or practical "good practices". Over time, however, the data documents a clear transition toward SoTL. Faculty members increasingly treat their classrooms as research environments, gathering empirical evidence on student performance, designing validated tools, and publishing their findings in international journals.
2. **Methodological Diversification and Active Learning (Nuñez–Cansado, 2025; UNESCO, 2026):** The content analysis tracks a systematic departure from traditional lecturing. Funded projects drove a diverse expansion of active learning models. Project–Based Learning, Challenge–Based Learning, clinical simulations, case methods, and educational Escape Rooms became standard practices. These approaches significantly enhanced student autonomy, teamwork, and critical thinking.
3. **From Basic Digitalization to the Artificial Intelligence Disruption:** Virtualization transformed dramatically over the eight–year timeline. Early projects focused on standard

Learning Management System (LMS) customization and basic digital resources (13% of the historical corpus). However, from the 2023/2024 cycle onward, the introduction of a dedicated institutional track sparked an exponential rise in Generative AI projects. This shift moved the institution away from simple digital setups and toward the ethical exploration of GenAI for personalized learning, creative assessment design, and curriculum restructuring.

4. **Structural Integration of Sustainability (SDGs):** The data shows that the SDGs evolved from being treated as optional, external syllabus topics into core structural elements of project design across all faculties. This allows students to connect academic theory with complex environmental and social challenges.

Conclusions

This eight–year longitudinal evaluation demonstrates that a structured, institutionally backed funding and evaluation framework successfully fosters a culture of continuous pedagogical excellence. Triangulating 146 approved projects and 114 final reports proves that competitive funding programs do more than improve immediate classroom dynamics; they actively promote long–term sustainability and drive systemic, university–wide transferability.

The study proves that individual teaching innovations can achieve lasting institutional value when wrapped in a clear quality assurance system and aligned with overarching university strategic plans. Furthermore, the rapid emergence of GenAI highlights that institutional innovation frameworks must remain agile. University policy should actively combine financial incentives with targeted pedagogical training. This dual approach empowers faculty to balance cutting–edge digital literacy with the ethical, human–centered responsibilities of modern higher education.

Keywords: Higher education; Teaching innovation projects; Longitudinal study; Evaluation

SESSION 3. INTERNATIONAL LARGE–SCALE ASSESSMENTS (ILSAS)
METHODS AND RESULTS
ORGANIZER: UNIVERSITY OF BATH
COORDINATORS: MARIA MAGDALENA ISAC – ANDRES SANDOVAL HERNANDEZ
5TH NOVEMBER: 2.00 P.M. – 4.00 P.M. {AULA A1.1 – RESEARCH 11}

Examining Environmental Behaviour Among Primary School Students Using TIMSS 2023: The Role of Environmental Attitudes and Parental Encouragement

Inés García Bohórquez – Giada Viotti – Camilo Ruiz – Laura Palmerio

Introduction

Education for Sustainable Development (ESD) has become a central priority in international and European education agendas in response to the growing environmental challenges (Calvin et al., 2023; The European Green Deal, 2019). Current policy frameworks emphasise that sustainability should be embedded across all aspects of school life, promoting not only environmental knowledge but also the competences needed to support sustainable development (Bianchi et al., 2022; UNESCO, 2017). Consequently, schools are increasingly expected to prepare students to become environmentally responsible citizens from an early age (Eurydice, 2024).

Achieving this objective requires a better understanding of the mechanisms through which environmentally responsible behaviours develop. Although environmental education has traditionally focused on strengthening students' environmental knowledge, research consistently shows that knowledge alone does not necessarily translate into environmentally responsible behaviour. This knowledge–behaviour gap indicates that students who understand environmental issues are not always more likely to engage in sustainable actions (Ballegeer et al., 2024; Oinonen et al., 2024; Portus et al., 2024). Identifying these mechanisms is therefore essential for informing both educational research and evidence–based policy (Lehre et al., 2026).

International Large–Scale Assessments (ILSAs) provide a unique opportunity to investigate these mechanisms by extending their focus beyond academic achievement to include sustainability–related constructs. Recent assessment cycles have incorporated measures of students' environmental knowledge, attitudes, behaviours, and learning contexts, allowing researchers to generate comparable cross–national evidence on the factors associated with environmental education (Hastedt et al., 2026; Isac et al., 2025). In particular, the TIMSS 2023 *Environmental Attitudes and Behaviours Framework* extends the concept of environmental awareness printroduced in TIMSS 2019 by integrating environmental knowledge, attitudes, behaviours, and home learning contexts into a coherent conceptual framework (Bookbinder et al., 2025; Mullis et al., 2021; Reynolds & Komakhidze, 2022).

Two complementary theoretical perspectives help explain the mechanisms underlying environmentally responsible behaviour. At the individual level, the Theory of Planned Behaviour (TPB) (Ajzen, 1991) proposes that behaviour is influenced primarily by attitudes rather than directly by knowledge. Accordingly, environmental attitudes are expected to mediate the relationship between environmental knowledge and environmentally responsible behaviour. At the contextual level, the Whole School Approach (WSA) conceptualises sustainability as a shared responsibility extending beyond classroom instruction to include schools, families, and the wider community (Verhelst et al., 2020, 2023). Within this perspective, students' environmental learning is shaped by supportive learning environments and parental involvement, both of which are considered

important for fostering sustainable behaviours (Wals et al., 2024). Despite the growing prominence of the WSA in educational policy, empirical evidence on how these contextual factors contribute to students' environmental outcomes remains limited, particularly in primary education and across different national contexts.

Building on these theoretical perspectives and recent developments in ILSAs, the present study integrates the TPB and the WSA using data from TIMSS 2023. Specifically, it examines whether environmental attitudes mediate the relationship between environmental knowledge and environmentally responsible behaviour while assessing the contribution of parental encouragement as a contextual factor within the WSA. By integrating individual and contextual perspectives, the study provides new cross–national evidence on the mechanisms underlying environmentally responsible behaviour among primary school students and further demonstrates the potential of TIMSS 2023 to advance research and policy on ESD.

Data

This study uses Grade 4 data from the TIMSS 2023 in eight European Education Area countries: Bulgaria, Cyprus, Ireland, Italy, Poland, Portugal, Slovakia, and Spain. Student and home questionnaire data were used to examine four constructs: environmental knowledge, environmental attitudes, environmentally responsible behaviour, and parental environmental encouragement. Environmental knowledge was measured using the TIMSS Environmental Awareness plausible values (ASSENV01–05), attitudes using the *Students Value Environmental Preservation* scale (ASBGVEP), behaviour as a latent construct based on six student items (ASBG12A–F), and parental encouragement as a latent construct based on five home questionnaire items (ASBH09A–E). The selection of these variables allowed us to examine both the individual mechanisms linking environmental knowledge, attitudes, and behaviour, and the contextual contribution of parental encouragement to students' environmentally responsible behaviour.

Methodology

All analyses were conducted in R using the *lavaan* package for confirmatory factor analysis and structural path modelling, *semTools* for psychometric analyses and *dplyr* for data preparation and filtering (Jorgensen et al., 2024; Rosseel, 2012; Wickham et al., 2026). The analytical strategy comprised two stages.

First, confirmatory factor analyses (CFAs) validated the latent constructs of environmentally responsible behaviour (six student items, ASBG12A–F) and parental environmental encouragement (five home items, ASBH09A–E) using WLSMV estimation with TIMSS sampling weights, given the ordinal nature of the indicators. Factor scores were then extracted from the validated models and used in the subsequent structural analyses. Second, structural equation modelling (SEM) examined relationships among environmental knowledge, environmental attitudes, parental environmental encouragement, and environmentally responsible behaviour. Environmental knowledge was represented using TIMSS plausible values (ASSENV01–05), attitudes by the ASBGVEP scale, and behaviour and parental encouragement by CFA–derived factor scores. The structural model was used to examine the direct effects of environmental knowledge and parental environmental encouragement on students' environmental attitudes and environmentally responsible behaviour, as well as the indirect effect of environmental knowledge on behaviour through environmental attitudes. Models were estimated using robust maximum likelihood (MLR) with full information maximum likelihood (FIML) to handle missing data and TIMSS sampling weights. To assess the consistency of the proposed theoretical framework across educational contexts, the model was estimated separately for each of the eight participating countries. Direct, indirect, and total effects were examined to evaluate the mediating role of environmental attitudes and the contribution of parental environmental encouragement (Fishbein et al., 2025; von Davier et al., 2024).

Results and discussion

The findings suggest that environmentally responsible behaviour among Grade 4 students is shaped by both contextual and individual factors, highlighting the complementary contributions of the WSA

and the TPB. While effect sizes varied across countries, the models revealed broadly consistent patterns across the participating education systems. The discussion therefore focuses on these general cross–national trends.

First, parental environmental encouragement was positively associated with both environmental attitudes and environmentally responsible behaviour in all countries. Given the age of primary school students, this result underscores that environmental learning extends beyond the classroom and is embedded in broader socialisation contexts, particularly the family. It provides empirical support for the WSA, which emphasises collaboration between schools, families, and communities in sustainability education. From a policy perspective, the findings suggest that ESD initiatives may be strengthened by actively involving parents alongside school–based interventions.

Second, environmental attitudes emerged as a key mechanism linking environmental knowledge and behaviour. Environmental knowledge was positively associated with environmental attitudes in seven of the eight countries, indicating that greater awareness of environmental issues is generally associated with stronger pro–environmental attitudes. In turn, environmental attitudes showed the strongest and most consistent association with environmentally responsible behaviour across countries. Moreover, attitudes mediated the relationship between knowledge and behaviour in nearly all countries, with Poland as the only exception. These results align with the TPB, which posits that attitudes are a central pathway through which knowledge influences behaviour. From an educational perspective, this suggests that increasing environmental knowledge alone may be insufficient to promote behaviour unless it is accompanied by the development of positive environmental attitudes.

A notable finding is that the direct association between environmental knowledge and behaviour becomes negative once environmental attitudes are included in the model. Rather than implying that knowledge undermines pro–environmental behaviour, this pattern suggests a more complex relationship. One possible explanation, discussed in recent environmental education literature, is that greater awareness of environmental problems may generate emotional responses such as environmental concern or eco–anxiety, particularly when not accompanied by perceived efficacy or opportunities for action. Although emotional processes were not examined in this study, they may help explain this pattern and represent an important direction for future research.

Conclusions

This study contributes to the ESD and ILSA literature by integrating the WSA and TPB to investigate the mechanisms underlying environmentally responsible behaviour among primary school students. Using TIMSS 2023 data from eight European Education Area countries, it demonstrates how theoretical models can be operationalised and tested in a cross–national context, highlighting the potential of ILSAs to advance research on sustainability education.

From an ESD perspective, the findings reinforce that environmental knowledge alone is unlikely to be sufficient to foster environmentally responsible behaviour. Instead, the development of positive environmental attitudes emerges as a central mechanism through which knowledge is associated with behaviour, consistent with an attitudinal process aligned with TPB. In addition, parental encouragement—operationalised as a proxy for the home dimension of the WSA—appears to play a consistent supportive role, highlighting the importance of linking classroom learning with family–based environmental socialisation. Together, these results underline the relevance of integrating school–based and home–based contexts in sustainability education.

From an ILSA perspective, the study demonstrates the potential of TIMSS 2023 to extend beyond cognitive achievement and by capturing meaningful variation in sustainability–related constructs across countries. It also highlights the importance of using validated measurement models and structural equation approaches to ensure comparability when analysing non–cognitive outcomes in international datasets.

Several limitations should be considered. The cross–sectional design does not allow causal interpretation, and the available indicators capture only part of the broader cognitive, contextual,

and affective dimensions relevant to environmental behaviour. In addition, while WSA is a multidimensional framework, its operationalisation here relies on available proxies within TIMSS, which may not fully capture its systemic nature. Nevertheless, the measurement approach provides a consistent and comparable approximation across countries. Future research should further explore affective and motivational factors such as emotions, perceived efficacy, and environmental concern to better understand how environmental knowledge is translated into behaviour. In addition, future cycles of PISA 2025, TIMSS 2027, and PISA 2029 will provide further opportunities to examine sustainability–related competencies across educational systems.

Keywords: Education for Sustainable Development; Theory of Planned Behaviour; Whole School Approach; TIMSS 2023

The effect of students' sport engagement at school on civic knowledge and social cohesion in Europe: Evidence from the ICCS 2022 survey

Marcos Delprato – Maria Magdalena Isac – Daniele Vidoni

Introduction

Although engagement in sport and physical activity in Europe has been stable and in line with 2017 pre–pandemic figures, the 2022 Eurobarometer dedicated to sport and physical activity shows that around 45% of Europeans aged 15–24 currently never exercise or partake in physical activity, a rate which is still 6 points larger than in 2009 (of 39%) (European Commission, 2022). Earlier literature on the relationship of practising sport with both cognitive and non–cognitive outcomes have gained significant attention on educational research (see the review of Barbosa, 2020; James et al., 2023; Jee, 2026). Frequent engagement in sport, both outside and within educational settings, can boost academic achievement and other social outcomes through its positive contribution to physical well–being and to cognitive and psychological development, which are then vital for academic success (Tomporowski and Pesce, 2019). There are different channels through which students' exercise can have leverage effects on academic performance (Roig–Hierro et al., 2025), namely: direct and indirect physiological, cognitive, and emotional (Santana et al., 2017). One of the main pathways is the neurophysiological one. Exercise has been demonstrated to increase cerebral blood flow which support brain plasticity and cognitive health, which are then related to higher cognitive capabilities (Cantelon and Giles, 2021) via the improvement of attention, working memory, and cognitive flexibility, all key elements for academic success (Vitiello and Greenfield, 2017).

However, as regards the connection of sport engagement of students within schools (school sport engagement: SSE) with civic outcomes, empirical evidence is nonexistent Some studies (e.g., Almeida et al., 2026) combine sport and civic and political activities, which are possible paths for positive youth development, but not modelling the impact of sport engagement on civic outcomes per se. In particular, based on ICCS data, there is a lack of evidence that extracurricular activities such as sport engagement in school can be a vehicle for boosting students' civic development and their degree of cooperation. Yet, community sports facilities funded by governments provide an inclusive environment where students from all backgrounds come together to engage in sports and physical activities, boosting diversity and belonging despite students' diverse backgrounds. The socialisation theory suggests that sports participation, as other youth activities, will raise the likelihood of prosocial outcomes (Rotolo et al., 2020). An important beneficial social outcome of practising sport at school is its socialising dimension, boosting students' empathy, teamwork, and tolerance (Zhou and Kaplanidou, 2018), as well as their democratic values and sense of shared responsibility (Stemhagen and Hytten, 2025).

Since the benefits of increasing sport engagement at school on students' civic outcomes (and communities' social cohesion) have not been quantitatively fully explored, this paper fills this empirical gap by assessing these two associations using the ICCS 2022 data. Moreover, from a policy angle, the paper's rationale fits within the European Sport Model vision of students' sport engagement as a pathway for increasing communities' social cohesion and boosting students social outcomes.

Data used

The paper relies on students and school data for the ICCS 2022 study resulting on a sample of around 50,000 observations from 17 EU countries (IEA, 2022). The student exercise indicator is defined at the school level. Specifically, from the school questionnaire, school principals were asked: 'During the current school year, how many [target grade] students in this school have had the opportunity to take part in any of these activities (sport events)', with answering being labelled as: 1 "None or hardly any", 2 " Some of them", 3 "Most of them", and 4 " All or nearly all".

Methodology

We use different models to estimate the association of SSE with students civic outcomes and community social cohesion. Initially, these associations are assessed without controls and then net of students and school controls. Specifically, using students' civic knowledge scores as dependent variable, two models are explored: one without controls (that is, only with the key binary explanatory variable SSE measuring sport engagement at school plus country fixed effects), and a full model (with additional controls: gender, language spoken and number of books at home, and index of ICT, school principal experience, civic education taught as separate subject, rate of disadvantaged students, and indices for community poverty and ethnic tensions). The models' focus is on the estimated impact of SSE on students' civic knowledge under these two modelling approaches (OLS for civic knowledge and logit to model the role of SSE in students reaching the top half of the civic knowledge distribution in each country). In addition, for community social cohesion, we present logit estimates of SSE on weakening community tensions due to ethnicity issues (also with and without controls). We are planning to extend our initial findings relying on additional modelling (e.g., quantile treatment effects, multilevel modelling, etc.)

Results and discussions

The paper's main findings are as follows. First, raw differences show that, for the whole sample of 17 countries, students who attend schools where most or nearly all of students are engaged in sport events in comparison to students who attend schools with weaker sport engagement (none or at most some students being involved in sport events) have, in average, a greater knowledge of civic outcomes (i.e., 523 versus 510), a statistically significant difference of 13 points. Whole sample's estimates show that the impact of attending a school with stronger students' engagement in sport events on students' civic knowledge is of +11 (and statistically significant). The relationship between SSE at school and civic knowledge is still present when accounting for drivers of civic outcomes and SSE in the full model. The impact is naturally of a smaller magnitude, going down from +11 to +5.4, though it is positive and statistically significant. Second, as regards to the analysis of drivers of low community's ethnic tensions, we find that SSE matters for inclusivity and social cohesion. That is, if students are fully engaged in practising in sports at school, this is related to weakening community tensions and raising inclusivity by 30% (estimated odds–ratio of 1.3 with full controls). Overall, results indicate that raising the extent of practising sports at school is a platform to improve students' civic outcomes as well as community social cohesion.

Conclusions

Using data from 17 EU countries from the ICCS 2022 data, this paper should be seen as an attempt to support the EU policy framework for sports. The paper fills a key empirical gap (i.e., sport and civic outcomes linkage) by offering new empirical evidence of students' engagement in sport at school on both students' civic knowledge levels and community social cohesion. The paper's findings also endorse earlier studies which emphasise the role that sports can accomplish for education, health,

well–being, integration/inclusiveness, and democratic values in the European region. Key findings are: (i) students' and schools' background play a significant role on the probability of students being engaged in sport events at school, with important difference within schools in each country; (ii) students attending schools where most of them are engaged in sport events show a positive gap for civic knowledge scores of 13 points in comparison to students attending schools with weaker sport engagement, with a positive gap still holds of 5.4 points when adding students' and schools' controls; and (iii) the chances of a community to have above average social cohesion can be increased by 30% under a scenario of strong school sport participation from students.

Keywords: Sport practice at school; Civic knowledge; Social cohesion; ICCS 2022

Distinguishing Substantive Attitude Profiles from Method Classes in the ICCS Students' Endorsement of Gender Equality Scale using ICCS

Natalia Lopez-Hornickel – David Torres Irribarra – Catalina Reszczynski – Diego Carrasco

Introduction

Gender attitudes are central to research on citizenship, democracy, and social cohesion. Brandt (2011), using the World Values Survey, provides evidence that sexist gender attitudes are associated with a decline in a country's gender equality over time, as measured by income gaps, political representation, and occupational opportunities. This is central to a global context that seeks to address gender discrimination in political representation, employment, and exposure to violence, among other areas (UN Women, 2023). In this sense, egalitarian gender attitudes contribute to the commitment to constructing democratic societies (Blaskó et al., 2019), where civic knowledge alone is insufficient and requires the incorporation of affective and behavioural dimensions (Kennedy, 2019).

The International Civic and Citizenship Education Study (ICCS) measures the endorsement of gender equality in ICCS using a set of questionnaire items, which are combined using item response theory models (Schulz et al., 2018). The resulting scores are reported as weighted likelihood estimates and subsequently transformed into an international reporting metric to aid interpretation. In ICCS 2016, scales were standardised to have a mean of 50 and a standard deviation of 10 across participating education systems that met the sampling requirements (Köhler et al., 2018; Schulz et al., 2018). Consequently, values indicate students' relative location within this reference distribution rather than the proportion of items endorsed. Higher scores indicate stronger support for egalitarian principles, whereas lower scores indicate weaker endorsement (Köhler et al., 2018).

Under this framework, the scale summarises the degree to which respondents align with principles of equal rights and opportunities for women and men, integrating information from multiple items into a single latent continuum (Schulz et al., 2018). Differences along the metric, therefore, correspond to gradations in the strength of commitment. Treating the variable as continuous implies that commitment can vary in intensity without distinguishing among possible configurations of response heterogeneity. However, students who obtain similar overall scores may still exhibit different response patterns.

Person–centred approaches, such as latent class analysis (LCA) and latent profile analysis (LPA), allow for the identification of unobserved heterogeneity by grouping individuals with similar response configurations across a set of indicators (Nylund–Gibson & Choi, 2018). Instead of assuming that differences are fully captured by variation along a single continuous dimension, mixture models assess whether the population is better represented as a collection of latent subgroups, each characterised by its own set of response probabilities (Nylund–Gibson & Choi, 2018).

In a study of Latin American adolescents using data from the ICCS 2016, López–Hornickel and Sandoval–Hernández (2023) identified four groups that were present across countries: fully egalitarian students, hesitant egalitarians, normative egalitarians, and political sexists. While some adolescents supported equality across all domains, others endorsed general principles but withdrew their support when items implied women' participation in politics or depicted male privileges in labour markets.

Although person–centred modelling offers advantages, its application may introduce response biases in scales composed of positively and negatively worded items (Ponce et al., 2022). Previous studies have dealt with the dilemma about the nature of wording effects (Ponce et al., 2022), trying to determine if they are purely artefactual, which means they are results of measurement bias because of the instrument (Marsh, 1996), or if they are substantive, which implies there is effectively a trait underlying the responses (Tomas & Oliver, 1999).

The *Normative* Egalitarians profile (López–Hornickel & Sandoval–Hernández, 2023) presents a substantive interpretation in a context where there seems to be a paradoxical adherence to gender equity due to cultural remnants such as egalitarian essentialism (Cotter et al., 2011). This means that there is support for women to develop social and public roles, as long as they fulfil what is believed to be their "natural" responsibility for caretaking at home (Cotter et al., 2011). However, similar response patterns may arise from wording effects, as differential responses to positively and negatively worded items can introduce systematic bias and generate spurious multidimensionality unrelated to the construct (Ponce et al., 2022). Therefore, it is necessary to assess whether this profile reflects a substantive attitudinal configuration or a methodological artefact.

Purpose and Research Questions

The present study asks three questions: (1) do the substantive classes identified in prior work survive once a confirmatory class representing polarity–insensitive responding is added to the model; (2) how large is this method–related class, and from which substantive classes are its members drawn; and (3) does explicitly modelling response behaviour change conclusions about the nature of population heterogeneity in this scale.

Conceptual Framework

In scales that combine positively and negatively worded items, respondents may respond differently to item polarity, generating systematic inconsistencies unrelated to the underlying construct (Ponce et al., 2022). These differences in response behaviour can introduce bias into measurement and produce patterns that do not reflect substantive attitudes.

Wording effects have been shown to induce spurious multidimensionality and distort model fit, particularly when models assume that all observed variation reflects the construct of interest (Marsh, 1996; Ponce et al., 2022). Under these conditions, classes identified in mixture models may reflect method–driven heterogeneity rather than substantive population structure.

The specific class under analysis, endorsing general equality items while rejecting specific "male privilege" items, can be explained by two distinct psychological approaches. The substantive account draws on ambivalent sexism theory (Glick & Fiske, 1996, 2011), in which hostile and benevolent sexism can coexist and correlate positively, and on the cultural frame of egalitarian essentialism, which supports women's public and professional participation conditional on their still fulfilling a "natural" caretaking role (Cotter et al., 2011a). The method account draws on satisficing theory (Krosnick, 1991): when answering carefully is effortful, respondents may default to agreeing with every item regardless of its direction. Woods (2006) showed by simulation that even moderate rates of careless responding to reverse–worded items (up to 30%) can turn a genuinely unidimensional scale into a spurious two–factor structure, and unmodelled response style has been shown to bias trait estimates most severely at the extremes of a distribution (Bolt & Newton, 2011) and to substantially reduce correct latent–class classification when ignored (Huang, 2016).

These two accounts are not merely alternative descriptions of the same pattern; they generate different, testable expectations. If the "Normative Egalitarians"–type profile reflects genuine

ambivalent sexism and egalitarian essentialism, it should persist once a method class is introduced into the model, and it should relate meaningfully to theoretically relevant predictors such as authoritarianism. If, instead, the profile is an artefact of satisficing, its prevalence should shrink or disappear once polarity–insensitive responding is modelled separately, and any remaining association with substantive predictors should be weak.

Accordingly, the analysis tests whether previously identified profiles persist once response processes linked to item polarity are explicitly modelled. This provides a basis for adjudicating between the two accounts rather than assuming either one a priori.

Data and Measure

We use the ICCS 2016 Latin American module (Chile, Colombia, Dominican Republic, Mexico, Peru; N = 25,319 students), applying the sampling weights supplied with the data.

The Students' endorsement of gender equality scale comprises six items, that were dichotomised: three positively worded (equal opportunities in government, equal rights "in every way," equal pay for equal work) and three negatively worded (women should stay out of politics; men should have priority for scarce jobs; men are better qualified for political leadership). Following López–Hornickel & Sandoval–Hernández (2023), all items are recoded so that 1 denotes the egalitarian response.

Analytical Strategy

Three models are compared throughout this section: **Model 1**, a four–class solution with country–specific item probabilities left free; **Model 2**, the same four–class solution with item probabilities constrained equal across the five countries (testing measurement invariance); and **Model 3**, a five–class confirmatory model that adds a fixed method class to Model 2, with only class sizes freely estimated.

All models were estimated in Latent GOLD. As a baseline, the four–class solution was re–estimated with country–specific item probabilities left free, recovering four classes matching the profiles reported by López–Hornickel & Sandoval–Hernández (2023). The Fully Egalitarian class (53.6%) endorsed all six items strongly ($P \geq .90$ throughout). The Normative Egalitarian class (23.3%) endorsed the general equality items near–universally but showed markedly weaker endorsement on items concerning job priority, equal pay, and political leadership—consistent with the pattern of broad support for equality principles coexisting with more traditional views on specific economic and leadership domains previously described for this profile. The Political Sexist class (9.3%) showed the lowest endorsement of the four classes overall, distinguished in particular by weaker rejection of restrictions on women's political participation. The Hesitant Egalitarian class (13.8%) endorsed most items strongly but showed a marked dip specifically on the item concerning job priority under conditions of scarcity.

Because the analysis pools five national samples, we then tested measurement invariance by constraining item probabilities equal across countries. The constrained (15 parameters) and unconstrained (39 parameters) models returned an identical log–likelihood and L^2, indicating the four–class structure is invariant across countries; a replication from an alternative random seed confirmed the solution is a global rather than local maximum.

This invariant four–class solution was carried forward unchanged into a five–class confirmatory model. A fifth class was added with its response probabilities fixed in advance rather than estimated from the data: a probability of 1 of the egalitarian response on the first three items and a probability of 0 of the egalitarian response on the remaining three items, a profile intended to represent respondents who answer without regard to the specific content of each item. Only the five class sizes — not the response probabilities of any class — were estimated from the data. Because the four substantive classes are held fixed, any shift in how many people are classified into each one, relative to the four–class model, can be attributed specifically to the addition of the method class, not to a redefinition of the substantive classes themselves.

Results

Table 1 summarises fit statistics for Models 1–3. Model 1 and Model 2 returned identical fit (L^2 = 726.97 in both), confirming invariance across countries. Compared with Model 2, the substantive baseline, Model 3 improved fit (L^2 fell to 660.93), and both information criteria favoured it, while the total bivariate residual also declined (34.95 to 29.06). This came without a meaningful loss of classification quality (entropy R^2 .724 vs. .723; classification error .116 vs. .129).

Table 1. Fit statistics for the four–class (country–free and country–invariant) and five–class confirmatory models.

Model	Npar	L^2 (df)	BIC (LL)	Entropy R^2	Class. error
4–class, country–free	39	726.97 (2242)	19,615	.724	.116
4–class, country–invariant	15	726.97 (2266)	19,410	.724	.116
5–class (+ method class)	20	660.93 (2261)	19,387	.723	.129

Source: own elaboration based on ICCS2016.

Model 3 assigned 3.7% of the pooled sample to the method class. Because this is a new fifth category, the four substantive classes necessarily shrank in relative size: Hesitant Egalitarian rose slightly, from 13.8% to 14.3%; Fully Egalitarian was essentially unchanged, from 53.6% to 53.7%; Political Sexist rose slightly, from 9.3% to 9.7%; and Normative Egalitarian dropped the most, from 23.3% to 18.7%.

To see where the Normative Egalitarian cases went, we compared each respondent's most likely (modal) class under Model 2 and under Model 3. Under Model 2, about 5,480 respondents were classified into the Normative Egalitarian class. Of those, about 1,026 (19%) were reclassified into the method class under Model 3, and a further 63 moved elsewhere. The Fully Egalitarian class, by contrast, lost only 18 of its roughly 13,880 members, and the two remaining classes were essentially unaffected. In other words, the drop in Normative Egalitarian's prevalence is not spread evenly across the sample; it is almost entirely explained by this one group of 1,026 reclassified respondents. A further pattern emerged when the reallocated cases were examined by country: all 1,026 respondents who moved from the Normative Egalitarian class into the method class were Mexican students. This does not call the class as a whole into question, since 81% of its members remain there, spread across all five countries; rather, the specific subgroup whose classification was ambiguous enough to shift is concentrated in one country. This matches the country–level diagnostics: Mexico's rate of low–confidence classification fell from .043 to .019 once the method class was added, the largest improvement of any country, while the Dominican Republic, which contributed no reallocated cases, kept an unchanged, elevated rate (.088). This concentration could reflect a measurement or administration issue specific to Mexico, or a genuine country–specific attitudinal pattern; the present data cannot distinguish between these, so we flag it as a priority for follow–up rather than resolve it here.

Discussion and Conclusions

Three of the four substantive profiles are essentially unaffected by the addition of the method class, and even the fourth, Normative Egalitarian, retains 81% of its members once the alternative is available. This is not consistent with the profile being mostly a response artefact; if it were, a much larger share of its members should have moved once a more fitting alternative was offered. It is consistent with a genuine attitudinal profile, one combining general support for equality principles with more traditional views in specific domains, as already described in prior work. At the same time,

the fact that some members did move, concentrated in a single country, shows that a method effect of this kind is present in the data and is worth checking for whenever this scale is used.
The concentration of the reallocated cases in a single country change what this finding can be used to argue. If polarity–insensitive responding were a stable, universal trait, its incidence should not depend on which country a respondent is from. Finding it entirely within one country instead points to two possibilities that this study cannot separate: a measurement or administration issue specific to that country's implementation of the survey, or a substantive attitudinal pattern that is more common there and happens to resemble the theoretically defined method profile. Either way, national differences in the prevalence of the Normative Egalitarian profile, including those reported in prior work, should not be read as straightforward evidence of cross–national attitude differences until this ambiguity is resolved.
These findings suggest two practical steps for research using latent class models on scales with mixed item wording. First, any theoretically fixed method class should be checked carefully against the verified polarity of each item, since a class built on a mistaken assumption about wording direction will not test what it claims to test. Second, before treating a small, method–like class as evidence of an individual response style, researchers should examine whether it is evenly distributed across relevant subgroups, particularly countries in multinational data; a concentrated pattern, as found here, is at least as consistent with a country–specific measurement issue as with an individual trait, and warrants further investigation before either interpretation is adopted.

Keywords: Gender equality attitudes; Wording effects; Latent class analysis; Adolescents

Opening the black box of socioeconomic inequalities in civic knowledge: A multilevel decomposition of the socioeconomic gap

Francisco Gatica – Jan-Philipp Wagner

Democratic societies depend on citizens who are able to understand political institutions, evaluate public issues critically, and participate effectively in democratic processes. Civic knowledge represents the cognitive dimension of these democratic competences and is consistently associated with democratic attitudes and civic participation and engagement later in life (Schulz et al., 2023; Torney–Purta, 2002).
For this reason, civic knowledge has become a central outcome of citizenship education and one of the three principal constructs assessed by the International Civic and citizenship Education Study (ICCS). The ICCS examines how young people are prepared to take their roles as citizens in society. Through its different cycles, it has shed light on changes in students' civic knowledge across societies. It also has become an important source to understand how civic education takes place across different educational systems (Malak–Minkiewicz & Torney–Purta, 2021).
The increasing policy attention devoted to citizenship education reflects a broader concern regarding democratic resilience. For example, the European Commission has recently proposed that by 2030 at least 85% of students should reach an adequate level of civic knowledge (European Commission, 2025). Yet current evidence suggests that many educational systems are struggling to achieve this target.
According to the last ICCS cycle, none of the participating countries recorded a statistically significant increase in civic knowledge compared with the previous ICCS cycle in 2016. On the contrary, students in the higher proficiency level have decreased between cycles. Although average levels of civic knowledge remain below desired benchmarks in many countries, an equally important concern is that civic knowledge levels are unequally distributed across students. Across ICCS countries, students

from more advantaged socioeconomic backgrounds consistently outperform their less advantaged peers (Schulz et al., 2023).

Since civic knowledge is largely acquired through formal education, schools occupy a central position in understanding how these inequalities emerge and evolve. Yet whether schools can overcome socioeconomic differences in students' backgrounds remains a matter of debate.

In the case for civic learning outcomes, the contribution of schools has traditionally been viewed with considerable scepticism. Jennings and Niemi's (1974) seminal work on adolescent political socialisation contributed to the view that schools play a role in civic and political learning, but that this role is often modest when compared with family and broader socialisation influences. Explanations for this view are partly rooted in theories of early political socialisation, which emphasise the role of parents in shaping children's political attitudes and behaviour (Jennings et al., 2009). Students from higher socioeconomic background are more likely to experience "information redundancy", therefore, schools may contribute to a smaller extent in their achievement. On the opposite side, students from lower socioeconomic backgrounds may benefit more from schools resulting in a compensatory effect (Miranda, 2025).

One influential line of research has addressed this question through the compensatory hypothesis, asking whether particular school characteristics benefit disadvantaged students more than advantaged students. Although evidence suggests that some school processes – such as an open classroom climate (Campbell, 2008) – may reduce socioeconomic inequalities, findings remain mixed across studies and contexts (Miranda, 2025).

If schools are sites where socioeconomic inequalities are reproduced or mitigated, then the compensation hypothesis represents one possible expectation regarding how those school processes operate. However, before asking whether schools compensate for socioeconomic disadvantage, it is first necessary to understand what explains the socioeconomic gap itself. This study therefore shifts the focus from testing compensatory effects to decomposing the gap into the student– and school–level characteristics that account for it. Therefore, this study focuses on the following three questions:

- RQ1: What is the magnitude of the socioeconomic gap in student civic knowledge across the participating countries, and to what extent is this gap attributable to within–school versus between–school factors?
- RQ2: How do specific student characteristics and school processes contribute to explaining the socioeconomic gap at both the within–school and between–school levels?
- RQ3: To what extent is the observed socioeconomic gap explained by differences in the distribution of characteristics (composition effects) versus differences in the educational returns to those characteristics (coefficient effects)?

It is important to clarify that this study does not test the compensatory hypothesis directly. Testing the compensation argument typically requires interaction models to assess whether the effects of school resources vary by student socioeconomic status. Instead, this study takes a complementary approach. Before asking whether schools compensate for socioeconomic disadvantage, it is a necessary logical precursor to understand what explains the socioeconomic gap itself.

Rather than evaluating interaction effects, this study shifts the focus to a multilevel Oaxaca (1973)–Blinder (1973) decomposition framework. This approach partitions the mean difference between socioeconomic groups into differences in the distribution of characteristics and differences in their returns. By identifying the specific student and school factors that account for the gap, this study offers a complementary perspective that establishes a clearer empirical baseline for how school processes operate across different educational systems.

Data

This study uses data from the International Civic and Citizenship Education Study (ICCS) 2022, conducted by the International Association for the Evaluation of Educational Achievement (IEA). The present analysis includes 20 participating countries for which the relevant variables were available

and comparable[6]. ICCS is national representative of students in 8th grade in the participating countries, and it employs a two–stage stratified sampling design. To account for this complex sampling design, the analysis applies the student, class, and school sampling weights provided by ICCS, following the procedures recommended in the ICCS technical documentation (Schulz et al., 2018, 2024).

Variables used in the study are summarised in table 1.

	Variable name	Description
Outcome	PV1CIV–PV5CIV Civic knowledge	Student civic knowledge computed using five plausible values. (mean: 500, SD: 100)
Student level variables	SD_GENDER	Student gender (0 = Boy, 1 = Girl)
	S_POLDISC	Students' discussion of political or social issues outside the school
	S_CITEFF	Students' citizenship self–efficacy
	S_OPDISC	Students' perceptions of openness in classroom discussions
	S_CIVLRN	Student reports about civic learning at school
	S_SCHPART	Students' participation in civic activities at school
School level variables	SES_SCHOOL	Classroom average SES
Comparison group	S_HISCED recoded	Highest parental education recoded into two categories: Low SES – High SES • Low SES (S_HISCED <= 3: neither parent completed tertiary education) • High SES (S_HISCED == 4: at least one parent with tertiary education)

Method

This study focuses on explaining the socioeconomic gap in civic knowledge. Drawing on International Civic and Citizenship Education Study (ICCS) 2022 data, it applies a multilevel Oaxaca–Blinder Decomposition (OBD) approach to disentangle which school characteristics and processes explain the gap. This allows to contribute to the question on how much of the observed SES gap is associated with unequal access to school resources and processes, and how much remains after accounting for those differences.

Originally elaborated by Oaxaca (1973) and Blinder (1973) to explain differences in wage means between two groups, this technic has been largely used in the field of education for the past 25 years to explain differences between groups, including gender (Fortin et al., 2015; Gevrek et al., 2020), school sector (Castro Aristizábal et al., 2017; Parvez & Laxminarayana, 2022; Romuald, 2023), indigenous status (Arteaga & Glewwe, 2019; Blanco, 2019), among others.

By decomposing the difference in means between two groups (*H* as high–SES group and *L* as low–SES group in our study), this approach partitions the gap into two components, one corresponding to the part of the gap that is explained by differences in the characteristics of the two groups under analysis (composition effect), and the part of the gap that is due to differences on the return of those characteristics (coefficient effect).

To avoid ecological fallacies this study uses multilevel approach acknowledging the nested structure of the data (Raudenbush & Bryk, 2002). By doing so, it also allows to differentiate if the gap is explained by differences in characteristics and they return in both within and between schools. The

[6] Countries included in the analysis: Bulgaria, Colombia, Cyprus, Denmark, Spain, Estonia, France, Croatia, Italy, Lithuania, Latvia, Malta, Netherlands, Poland, Romania, Serbia, Slovakia, Slovenia, Sweden, Taiwan

former stands for the part of the gap that is explained by differences between students in the same class, whereas the latter, stands for the share of the gap that is due to differences between schools. Combining the within– and between–school differences together with composition and coefficient effect, this approach allows to identify four components summarised as follow:

	Composition effect	Coefficient effect
Within–School	Differences in individual characteristics Part of the gap due to differences in student characteristics and resources within the same school	Different slopes within schools Part of the gap due to differences of the return of student's characteristics and resources within the same school
Between–School	Differences in school averages Part of the gap due to differences between schools in student composition	Different contextual slopes Part of the gap due to differences of the return of the school average of students' characteristics and resources

To operationalize the decomposition described above, a custom analytical pipeline was developed in R as no existing software package currently implements an OBD within a multilevel modelling framework. Multilevel linear models were estimated separately for each socioeconomic group using the WeMix package in R (Sikali et al., 2023). Following Rubin's (1987) rules, the point estimates of the decomposition components were averaged across the five imputations.

Results

The civic knowledge gap between low– and high–SES students is statistically significant across all 20 participating countries, ranging from approximately 25 to 65 points on the ICCS scale. Applying the multilevel decomposition reveals that this gap reflects both differences in the distribution of student and school characteristics (composition effects) and differences in their educational returns (return effects).

The composition component is driven primarily by differences between schools rather than differences between students attending the same school. Across almost all countries, within–school composition contributes only marginally to the socioeconomic gap (typically around 2–3 points), whereas school–level composition accounts for the largest share of the explained gap. This indicates that socioeconomic inequalities in civic knowledge are strongly associated with the characteristics of the schools attended by students rather than with differences among students within the same school.

Regarding individual variable contribution, two findings stand out. On the one hand, and aligned with the literature, differences in open classroom discussion between low– and high–SES students explain to some extent the gap at both within– and between school level. Within school differences are very modest explaining up to 2 points of the civic knowledge scale, but consistently across countries. Between–school differences, explain a larger amount of the gap (up to 12 civic knowledge points) although only in half of the countries. On the other hand, within–school differences on student's participation in civic activities at the school explain up to 3 points of the civic knowledge scale, but only in half of the countries.

Return effects are dominated by the intercept term, while the estimated return effects associated with observed student and school characteristics are generally small and rarely statistically significant. This suggests that differences in the educational returns to observed characteristics are largely associated with factors not captured by the current model, therefore, differences in observed characteristics alone are insufficient to account for the socioeconomic gap.

Looking across the decomposition, the largest composition effects are consistently associated with contextual school characteristics, particularly school socioeconomic composition, whereas individual student characteristics contribute comparatively little. This suggests that inequalities in

civic knowledge are shaped less by differences among students within the same school than by the educational contexts in which they are enrolled.
These findings offer valuable contributions to international educational research and policy, with clear implications for education systems. By demonstrating that the socioeconomic gap in civic knowledge is primarily driven by between–school stratification rather than within–school differences, the study indicates that individualised student interventions alone are insufficient. For Southern European educational systems like Italy, which often navigate regional socioeconomic variations and structural disparities, these results suggest that policy interventions must target institutional equity and the socioeconomic composition of schools.
In terms of practice, the consistent evidence regarding open classroom discussions demonstrates that fostering an open environment for democratic dialogue is a practical and scalable school–level process that can mitigate inequalities. Finally, for research and evaluation bodies (e.g., INVALSI), this study establishes a robust analytical baseline. It moves the academic conversation beyond interaction models to isolate the precise structural drivers that educational systems must address to meet the European Commission 2030 benchmarks.

Keywords: Civic Knowledge; Socioeconomic Inequalities; Citizenship Education; Gap Decomposition

Professional competences and reflexivity in support teachers: evidence from the TFA-INDIRE post-training questionnaire

Sara Mori – Serena Greco – Francesa Storai – Elsa Maria Bruni

Introduction

The professional profile of the support teacher should not be reduced to a specialist function centred only on pupils with disabilities. In the Italian school system, the support teacher has a specific mandate, but this mandate is embedded in the broader professional community of teachers and in the shared responsibility for inclusive education. The Ministerial Decree of 30 September 2011 defines the support teacher through a complex set of pedagogical, didactic, psychological, relational and organisational competences (Italian Ministry of Education, University and Research, 2011). This contribution aims to investigate how teachers enrolled in TFA-INDIRE support teacher training perceive their own competences at the end of the course. The focus is not on disability as an isolated object of intervention, but on the competence profile that allows the support teacher to act within classroom planning, collegial work, relations with families, school organisation and local networks. This perspective is consistent with a view of inclusive education as a transformation of ordinary teaching and school cultures, rather than as a separate remedial track (Florian, 2014). The theoretical framework also refers to literature on teacher professionalism. The support teacher is considered a full member of the teaching profession, whose specialisation qualifies and extends ordinary teaching responsibilities. In this sense, specialisation does not mean separation from the class group, but the capacity to make inclusive knowledge available to the whole school community. This interpretation is coherent with the idea of inclusion as a general pedagogical task, not only as a field of special intervention (Bertagna, 2023). The study gives particular attention to perceived competence. Self-perception is not treated as a weak substitute for observed practice, but as a relevant professional construct. Bandura's theory of self-efficacy shows that beliefs about one's own competence influence motivation, persistence and professional action (Bandura, 1997). For this reason, the analysis of perceived competences can offer useful evidence on how teachers position themselves in relation to the role of support teacher after a structured training pathway.
Professional reflexivity is considered a transversal competence that gives structure to these different functions. The notion of reflective practitioner is used to interpret the teacher's capacity to examine

situations, make educational decisions explicit and transform experience into professional knowledge (Schon, 1983). The formative value of reflection is also linked to the possibility of revising interpretative frames and professional identity (Mezirow, 1991).
This contribution develops a further analytical step from the previous empirical study by Bruni, Greco and Mori, published in IUL Research. That article analysed the same TFA-INDIRE training context and highlighted the centrality of professional reflexivity in explaining inclusive orientations among teachers (Bruni et al., 2026). The present proposal moves from that evidence to a more specific focus on the perceived competences of support teachers and on their organisational role in inclusive processes.

Data used

The empirical basis of the contribution is the post-training questionnaire developed for the TFA-INDIRE programmes for the specialisation in support teaching. The questionnaire was designed to collect participants' final observations on the training pathway, their perception of the support teacher role, and their perceived competence in relation to school inclusion.
The reference population includes 9,354 trainees enrolled in the TFA-INDIRE pathways activated under Ministerial Decrees 75 and 77/2025. The effective sample includes 3,760 respondents, with an overall response rate of 40.2%. This response rate is adequate for survey research on large professional populations (Nulty, 2008).
The sample includes teachers from pre-primary, primary, lower secondary and upper secondary school. Female teachers represent 77.8% of respondents, while male teachers represent 22.2%. The mean age is 47.6 years. The sample also covers all Italian macro-areas, allowing comparative analyses by geographical area, school level and training pathway.
The questionnaire is structured into four sections. The first section includes nine items on opinions about the training received. The second section includes twenty-one items on the perceived impact of training on the role of the support teacher. The third section includes nineteen items on perceived competences in key areas for inclusion. The fourth section contains open-ended questions that may support future qualitative integration.
The scale is a four-point Likert scale, from 1 = not at all agree to 4 = fully agree. The use of a Likert scale is appropriate for measuring attitudes and perceived competences in educational research (Likert, 1932). The absence of a neutral midpoint is intended to limit central tendency bias and to encourage respondents to express a clear position.

Methodology

The proposed analysis is quantitative and exploratory. It is designed to build a competence profile of support teachers by reorganising the questionnaire items according to the professional areas implied by the 2011 ministerial profile and by the structure of the post-training questionnaire. The approach is evidence-informed, because it uses empirical data to refine educational interpretation without reducing professional competence to purely technical indicators (Biesta, 2010).
First, descriptive statistics will be calculated for all questionnaire items. Means, standard deviations, medians and response distributions will identify the areas in which trainees report stronger or weaker perceived competences. This first step will make it possible to distinguish competences perceived as consolidated from competences that may require further training support. Field's methodological framework will be used to guide the treatment of ordinal survey data and the choice of appropriate statistical procedures (Field, 2013).
Second, the items will be grouped into theoretically defined competence areas. The proposed areas are: pedagogical-didactic competences; psycho-relational competences; observational and documentation competences; digital and assistive technology competences; assessment and feedback competences; collegial and organizational competences; family and territorial-network competences; ethical-deontological competences; reflective and professional-development

competences. This grouping follows the questionnaire structure and translates the support teacher profile into empirically analyzable dimensions.
Third, reliability analyses will be conducted for each competence area. Internal consistency indices will support the construction of synthetic competence scores where appropriate. This procedure is coherent with the use of self-report instruments for latent constructs such as professional attitudes, perceived role and self-assessed competences (Corbetta, 1999).
Fourth, an exploratory factor analysis will be conducted to verify whether the empirical structure of the data corresponds to the theoretically defined competence areas or whether broader latent dimensions emerge. Sampling adequacy will be checked through the Kaiser-Meyer-Olkin index (Kaiser, 1974). The analysis will use oblique rotation if the factors are expected to be correlated, as is plausible in the case of professional competences.
Fifth, regression models will be used to test the role of professional reflexivity. Reflexivity will be operationalized through the items on critical reflection on representations of disability and inclusion, reflection on educational practice, and use of pedagogical reflection and documentation for continuous improvement. The models will test whether reflexivity predicts perceived competences after controlling for demographic and professional variables.

Expected results and discussion

The analysis is expected to provide a multidimensional profile of support teachers' perceived competences at the end of the TFA-INDIRE training pathway. Rather than treating perceived competence as a single global construct, the study will identify different areas of professional functioning. This is particularly relevant because the support teacher role operates across several levels: the classroom, the teaching team, the school organization and the territorial network.
At the classroom level, the questionnaire is expected to capture competences related to inclusive and accessible lesson planning, active and collaborative methodologies, group management, accessible communication, assessment adaptation and feedback. These competences correspond to the ordinary teaching dimension of the support teacher's work and show how specialization is enacted in daily pedagogical decisions.
At the organizational level, the expected results should highlight the relevance of items concerning participation in school-level moments of reflection and planning, contribution to inclusion strategies, involvement in projects and relations with the school leader. This level is important for interpreting the support teacher as an organizational resource and not only as an individual specialist.
At the network level, the questionnaire is expected to show the perceived capacity to interact with families, health services, social services, associations, cooperatives and local authorities. These competences are closely connected to continuity, shared planning and the construction of inclusive trajectories beyond the classroom.
At the reflective level, the expected findings should clarify whether reflexivity acts as a transversal competence. The working hypothesis is that teachers who report higher levels of critical reflection and documentation also report higher perceived competences in didactic, relational and organizational areas. This would confirm the role of reflection as a professional mechanism that connects training, practice and improvement.

Conclusions

The contribution proposes a re-reading of the TFA-INDIRE post-training questionnaire as a tool for analyzing the perceived professional competences of support teachers. The focus is placed on competence perception in continuity with the 2011 ministerial profile, with particular attention to the pedagogical, relational, organizational and reflective dimensions of the role. This allows the questionnaire to be used not only as a satisfaction or evaluation tool, but also as a research instrument for describing support teacher professionalism.

The main expected contribution is twofold. First, the study can provide a quantitative mapping of the competences that support teachers perceive as acquired or strengthened at the end of the training pathway. Second, it can test whether professional reflexivity functions as a key predictor of perceived competence. This hypothesis is coherent with the idea that teacher education should form professionals able to interpret, document and improve their practice, rather than simply apply procedures (Schon, 1987).

From this perspective, the support teacher is not only a specialist for disability. The support teacher is a reflective professional and an organizational resource for inclusive education. Training should therefore strengthen not only individual intervention skills, but also co-planning, documentation, collaboration, participation in school decision-making and networking with families and services. The adult-learning perspective is useful here, because it reminds us that professional development is stronger when it is connected to experience, agency and situated problems of practice (Knowles et al., 2015).

The previous IUL Research article showed the relevance of reflexivity within the same TFA-INDIRE training context (Bruni et al., 2026). This new analytical focus can extend that result by showing how reflexivity is connected to specific areas of perceived competence. In this sense, the study can contribute to a more operational understanding of support teacher professionalism and to the design of future training pathways.

The study has some limits. The data are self-reported and concern perceived competences, not directly observed practices. For this reason, future developments should integrate questionnaire data with qualitative analyses of open-ended responses, interviews, focus groups and classroom observations. Attention should also be given to common method bias, which can affect self-report surveys (Podsakoff et al., 2003).

Despite these limits, the proposed analysis can offer useful evidence for training design and for school improvement. It can help identify the areas in which support teachers feel more or less competent, the organisational conditions that support their role, and the reflective processes that make inclusion a shared professional responsibility.

Keywords: support teachers; perceived professional competences; professional reflexivity; inclusive education

SESSION 9. DIGITAL COMPETENCE AND GENERATIVE AI IN EDUCATION: NEW CHALLENGES FOR TEACHING AND ASSESSMENT

ORGANIZERS: INVALSI – UNIVERSITY OF BERGAMO

COORDINATORS: PAOLO BARABANTI – MARCO GIGANTI

5TH NOVEMBER: 2.00 P.M. – 4.00 P.M. {AULA B0.1 – RESEARCH 12}

Generative AI Chatbots in Classroom Management Activities in Higher Education: A Qualitative Study Based on Third–Generation Activity Theory

Stefano Bonometti – Letizia Ferri

Introduction

The growing integration of generative artificial intelligence in higher education is transforming the ways in which students approach learning, collaboration and complex educational situations. In the field of continuing education and workplace learning, AI–based chatbots can support students in analysing problems, generating ideas, structuring possible responses and reflecting on classroom management choices. However, their use also raises important questions concerning participation, relational dynamics, the distribution of cognitive work and the perception of AI as a meaningful interlocutor.

This study investigates the impact of generative AI chatbots on classroom management activities in the university context. The theoretical framework is grounded in third–generation Activity Theory, which makes it possible to interpret AI not merely as a technical tool, but as a co–actor within an activity system. From this perspective, the introduction of AI may reshape the relations among subjects, tools, rules, community and division of labour, generating new forms of collaborative learning as well as potential transformative tensions.

Data Used

The research was conducted within the course Methodology of Continuing Education in Work Contexts during the academic year 2025–2026. The experimental context involved approximately twenty university students engaged in the analysis and management of complex educational and classroom situations. The empirical material consisted mainly of qualitative data collected through focus groups, observations and the analysis of recorded and transcribed interactions. These data enabled an in–depth exploration of students' collaborative dynamics, their use of AI during classroom management activities and their perceptions of the relationship with the chatbot.

Methodology

The study adopted a mainly qualitative research design. The experimentation was organised into three progressive phases. In the first phase, AI was used primarily as a consultation tool. In the following phases, its role gradually evolved, becoming increasingly integrated into the students' activities and eventually functioning as an active partner in the discussion and interpretation of classroom management situations. Focus groups were used to collect students' reflections on the experience, with particular attention to three dimensions: the influence of AI on problem solving, the level of participation within working groups and the perception of empathy and relational proximity in the interaction with AI. Observations and transcribed interactions were analysed in order to identify how students negotiated meanings, distributed tasks, used AI outputs and redefined their own roles within the collaborative process.

The analytical lens of Activity Theory supported the interpretation of the chatbot as an element capable of modifying the structure of the activity system. Particular attention was paid to the tensions

emerging between human agency and AI support, between individual and collective decision–making, and between instrumental use and relational attribution.

Results and Discussion

The study highlights that the integration of generative AI influenced the students' approach to problem solving by expanding the range of possible interpretations, supporting the formulation of hypotheses and stimulating reflection on the coherence of classroom management choices. AI appeared to function as a cognitive and dialogical resource, especially when students used it not only to obtain answers, but also to test ideas, compare alternatives and refine their reasoning.

The introduction of AI also affected participation within working groups. In some cases, the chatbot supported more distributed forms of contribution, offering shared material around which students could discuss, negotiate and make decisions. At the same time, the presence of AI required students to redefine their roles and responsibilities, particularly in relation to the evaluation, selection and transformation of AI–generated suggestions.

A further relevant dimension concerns the perception of empathy and relationship with AI. Although the chatbot does not replace human interaction, students tend to attribute a dialogical function to it while at the same time expressing caution and a sense of disorientation regarding direct dialogue with the chatbot. This suggests that the relational dimension of AI use cannot be reduced to technical efficiency alone; rather, it should be understood within a broader horizon that is nevertheless not perceived as human.

From the perspective of third–generation Activity Theory, these findings show that AI can generate changes in the relationships among subjects, tools, rules, community and division of labour. The chatbot becomes a mediating artefact that participates in the reorganisation of classroom management activities, producing both opportunities for learning and tensions that require critical reflection.

Conclusions

The research suggests that generative AI chatbots can play a significant role in supporting students during classroom management activities in higher education. Their value does not lie only in the production of content, but also in their capacity to stimulate problem solving, support collaborative discussion and contribute to the reconfiguration of roles within group work.

At the same time, the study underlines the importance of adopting a critical and pedagogically grounded approach to AI integration. Students need to learn how to question, interpret and transform AI outputs rather than accept them passively. In this sense, AI can become a meaningful component of educational processes when it is embedded in reflective, collaborative and theoretically informed learning environments.

The use of third–generation Activity Theory offers a robust framework for understanding these transformations. It allows researchers and educators to analyse AI as part of a complex system of relations, practices and tensions, rather than as an isolated technological instrument. Future developments of the research may further examine how different modes of AI integration affect participation, agency and learning outcomes in university–based classroom management activities.

Keywords: Activity Theory; Chatbot; High Education; Classroom

Mapping the Diffusion of Artificial Intelligence in Italian Schools: A Text Mining Analysis of 2022–2025 and 2025–2028 PTOF Documents

Giovanni Piumatti

Artificial intelligence (AI) is considered a major driver of change in education systems. Schools are expected to prepare students for labour markets transformed by automation, data–intensive work and generative AI, while ensuring that new technologies do not exacerbate existing inequalities in learning opportunities. These challenges are especially relevant in Italy, where school digitalisation intersects with long–standing territorial disparities, changing skills needs, and the need to support teachers and school leaders in integrating new technologies into educational practice. Systematic evidence on how Italian schools address AI at the institutional level remains nevertheless limited. , Existing surveys mostly concerns teachers' individual use of AI, such as the OECD TALIS 2024, which reports that 25% of Italian lower–secondary teachers had used AI in their work at least once in the previous year, compared with an OECD average of 36%. Much less is known about whether and how schools integrate AI into their formal planning. This matters because AI adoption is not only about teachers' individual practices or technological availability; it also concerns how schools interpret AI, connect it to their educational mission, plan training and innovation strategies, and frame it as a pedagogical opportunity, labour–market competence, ethical challenge or component of digital transformation.

This study addresses this gap by analysing the Piani Triennali dell'Offerta Formativa (PTOF), the three–year strategic and pedagogical planning documents that every Italian school must prepare and submit to the Ministry of Education. PTOFs describe schools' educational priorities, their curricular and extracurricular offerings, their organizational structure, their innovation strategies, and their medium–term planning, thereby providing a valuable resource for examining how schools position themselves regarding new educational innovations. Building on previous analyses of the 2022–2025 PTOFs, this study compares two consecutive planning cycles, 2022–2025 and 2025–2028, to assess whether AI has moved from an emerging topic to a more central component of school planning and in which areas.

The study makes two contributions. First, it provides a descriptive and comparative analysis of the institutional visibility of AI references in Italian schools over time. Second, it lays the groundwork for a text–mining analysis examining not only whether schools mention AI, but how they frame it. The overarching research question is: how has the presence of AI in Italian schools' strategic planning changed between the 2022–2025 and 2025–2028 PTOF cycles, and what does this reveal about the institutionalisation of AI in the Italian education system? More specifically, the study asks: how has the proportion of retrieved PTOFs mentioning artificial intelligence changed across the two cycles? What territorial differences emerge, and how do they evolve? How do AI references vary by school level and type? Finally, how is AI framed in school planning documents, and how do these semantic framings change over time? To answer these questions, we created a harmonised database of PTOF documents downloaded from the Ministry of Education's official archive, Scuola in Chiaro. The analyses cover public and private schools from primary to upper–secondary education, excluding pre–schools. Documents were converted into machine–readable text and searched for the keyword "intelligenza artificiale". The resulting dataset was linked to administrative information on school location and type. The analysis focuses on explicit references to AI; it does not measure actual classroom use or implementation quality, but rather the institutional visibility of AI in official planning documents.

Document retrieval was geographically uneven. For 2022–2025, 8,564 PTOFs were retrieved from 11,808 institutions in July 2025, corresponding to a national retrieval rate of 73%. Excluding Bolzano and Trento, where PTOFs were not available online, average provincial coverage was 78%, ranging from 55% in Ascoli Piceno to 95% in Brindisi and Mantua. For 2025–2028, 8,566 PTOFs were

retrieved from 11,686 institutions in July 2026, again corresponding to a national retrieval rate of 73%. Average provincial coverage rose to 80%, ranging from 28% in Parma to 96% in Piacenza and Pisa. Overall, 11,613 school records appear in both cycles, allowing the analysis to capture broader institutional change rather than two unrelated cross–sections.

In the 2022–2025 PTOFs, 3,695 of the retrieved documents mentioned AI, equal to 43% of retrieved PTOFs. At provincial level, the average share was 45%, ranging from 23% in Verbano–Cusio–Ossola to 73% in Teramo. In 2025–2028, the number rose to 6,261, equal to 73%. The average provincial share increased to 75%, ranging from 43% in Parma to 92% in Matera. These findings indicate a rapid increase in the institutional visibility of AI: within one planning cycle, AI moved from being mentioned in fewer than half of retrieved PTOFs to almost three–quarters of them.

Territorial differences were relevant. In 2022–2025, previous analyses identified a north–south gradient, with schools in central and southern provinces more likely to mention AI than those in northern provinces. In 2025–2028 this pattern seemed to persist, despite strong growth everywhere. The average share of retrieved PTOFs mentioning AI rose from 40% to 71% in the North, from 46% to 75% in the Centre, and from 51% to 79% in the South and Islands. Thus, AI visibility is increasing across the system, but territorial convergence remains incomplete. Differences by school level and type were also significant. In 2022–2025, AI references were more frequent in upper–secondary schools, especially academic followed by technical–vocational tracks, than in lower–secondary and primary schools or private institutions. This hierarchy remained visible in 2025–2028, although all categories showed substantial growth. Among retrieved PTOFs, academic tracks increased from 72% mentioning AI to 90%, and technical–vocational schools from 65% to 89%. AI has therefore become almost standard in upper–secondary planning. Growth was also substantial at earlier school levels: primary schools increased from 29% to 65% and lower–secondary schools from 53% to 67%. This suggests that AI is no longer confined to upper–secondary education but is increasingly entering the planning documents of schools serving younger pupils. On the other hand, private schools remained distinct. Although AI mentions increased in private institutions, their levels remained lower than in most public school categories. Private primary schools rose from 4% to 19%, private lower–secondary schools from 12% to 34%, and private upper–secondary schools from 16% to 38%. This indicates that the spread of AI references was broad but uneven, with public upper–secondary schools leading the integration of AI into formal planning.

These findings suggest that AI has rapidly become part of the common vocabulary of school planning in Italy. The increase from 43% to 73% of retrieved PTOFs reflects a major shift in how schools position themselves toward digital innovation and future skills. However, the persistence of territorial and institutional differences shows that diffusion does not imply convergence. The strong presence of AI in upper–secondary schools suggests a connection with digital professions and labour–market preparation, while its growing presence in primary and comprehensive schools may indicate increasing relevance for digital citizenship, foundational competences and responsible technology use.

The presence of the keyword alone does not show how schools interpret AI, what purposes they associate with it, or whether they frame it as an opportunity, a risk, a competence area, an ethical issue, a teaching aid or an organisational challenge. Accordingly, the next phase will extend the analyses through text mining and semantic classification. Drawing on the full corpus of PTOFs from 2022–2025 and 2025–2028, the analyses will combine quantitative natural language processing techniques with human judgement through semantically defined categories. The text–mining component will involve document cleaning, normalisation and indexing; keyword–in–context and collocation analyses; extraction and deduplication of passages surrounding AI mentions; and classification into interpretable semantic categories. These categories will capture the different roles assigned to artificial intelligence in school planning, distinguishing between pedagogical uses, student competence development, integration into STEM and digital curricula, ethical and regulatory issues, teacher–training needs, and broader school innovation strategies. By combining descriptive

and semantic analyses, the study will provide a more comprehensive account of the institutional uptake of AI in Italian schools. In sum, while preliminary quantitative evidence showed that AI has moved rapidly from emergence to mainstream visibility in PTOF documents, the forthcoming qualitative and computational analysis will assess whether this growth reflects a substantive transformation in schools' strategic narratives, or whether AI is often incorporated through generic references to digital innovation.

Keywords: Artificial intelligence; School planning; Text mining; PTOF

Generative Artificial Intelligence and Universal Design for Learning: Evidence from Literature and from an Exploratory Qualitative Study with Teachers

Antonella Coppi – Sabrina Lucilla Barone – Debora Tringali

Introduction

The rapid dissemination of Generative Artificial Intelligence (AI) is progressively reshaping teaching, instructional design, assessment practices, and forms of educational participation. At the same time, Universal Design for Learning (UDL) provides a pedagogical framework aimed at reducing barriers to learning and promoting multiple means of engagement, representation, action, and expression (CAST, 2024). Within this scenario, the relationship between Generative AI and UDL assumes particular relevance for rethinking inclusive teaching practices, teachers' digital competence, and formative assessment.

Digital competence cannot be reduced to mere technical mastery; rather, it must be understood as a critical, ethical, reflective, and pedagogically situated competence (Vuorikari et al., 2022). In the same direction, the most recent international frameworks emphasize the need to develop teachers' AI literacy while preserving human agency, professional autonomy, and educational responsibility (European Commission, 2022; UNESCO, 2024). From this perspective, the formalization of an "algorethical" theory becomes necessary to guide critical and ethical approaches that integrate AI into curricula while preserving human values and deconstructing the myths of technological hypertrophy and speed (Perla et al., 2024). To this end, the integration of algorithmic systems into education requires a systemic and transdisciplinary approach, rooted in planetary humanism, capable of transforming AI from a merely technical construct into a catalyst for ethical responsibility and social justice (Carletti, 2025).

A scoping review of the international literature, updated to June 2026 and focused on the integration of Generative AI and UDL in teachers' professional practices, reveals a still recent and fragmented field of research, in which conceptual contributions, small–scale empirical studies, and exploratory research prevail. The review considered inclusive education, teachers' digital competence, accessibility, formative feedback, and assessment, identifying four recurring lines of inquiry.

The first concerns co–design: AI is used to adapt lesson plans, differentiate materials, generate scaffolds, and prepare multimodal resources, although the outputs do not automatically meet UDL criteria; therefore, critical selection, revision, and contextualization by the teacher remain necessary (Evmenova et al., 2024; Barbieri et al., 2026). The second concerns accessibility, through text simplification and reformulation, multilingual translation, speech–to–text and text–to–speech technologies, subtitling, alternative texts for images, and personalized supports. The third addresses feedback and assessment: speed and personalization appear promising, but the formative value of feedback depends on its dialogic, situated, and pedagogically oriented nature (Huesca et al., 2025). In this sense, the integration of AI's analytical and adaptive potential with the Universal Design for Learning (UDL) framework makes it possible to structure inclusive environments capable of offering multiple representations and tailored pathways from early childhood onward, fostering students'

agency and responsibility (Baldassarre, Ruotolo, & Sarcina, 2025). The fourth concerns AI–based assistive technologies, for which the initial evidence suggests possible effects on the engagement and outcomes of students with disabilities, while also requiring further controlled and longitudinal investigation (Anwar et al., 2026). The available evidence therefore outlines a significant educational potential, the realization of which depends on teachers' ability to evaluate outputs, recognize barriers, protect students' agency, and maintain responsibility for instructional and assessment decisions. Less explored, however, is the way in which these needs are interpreted in school contexts and translated into professional requirements. This problematic space provides the basis for an exploratory qualitative study, currently underway, involving upper secondary school teachers. The study is guided by the following research questions: RQ1. What representations of Generative AI emerge from teachers' experiences and reflections? RQ2. What opportunities and critical issues are associated with its use in heterogeneous classrooms and inclusive processes? RQ3. What competences and forms of professional support are considered necessary for an inclusive, critical, and pedagogically responsible use of AI?

Data used

The empirical corpus consists of responses collected through an online questionnaire addressed to upper secondary school teachers. At the cut–off point defined for the present analysis, 33 completed questionnaires were available. Participants work mainly in the Campania region and teach in different types of upper secondary schools, including academic–track high schools, technical institutes, and vocational schools, across a range of disciplinary areas. The group includes 29 women and 4 men, aged between 30 and 67, with diversified professional experience profiles.

The data include socio–professional information, short questions on previous experiences with AI and UDL, and a textual corpus produced through open–ended questions concerning representations of AI, opportunities and critical issues in inclusive contexts, training needs, and forms of professional support. Data collection is ongoing; the results presented refer exclusively to the 33 completed questionnaires included in the frozen corpus for analysis and should be considered preliminary.

Methodology

The study adopts an exploratory qualitative design with an interpretive orientation, aimed at understanding the meanings, beliefs, cautions, and training needs that teachers associate with the use of Generative AI in inclusive school contexts. The investigation does not aim to estimate the prevalence of these phenomena within the teaching population, but rather to reconstruct participants' representations and generate interpretive categories from the textual material (Trinchero, 2004). Data were collected through a self–administered online questionnaire, specifically designed for the study and administered via Google Forms. Participation was voluntary. Before administration, participants were informed about the aims of the research, the anonymous nature of the survey, and the ways in which the data would be used; participation in the study and completion of the questionnaire constituted consent to take part. The instrument comprises 16 items, in addition to the automatic recording of the time of completion and is organized into four areas. The first collects five socio–professional details: age, gender, years of service, type of school, and subject taught. The second includes four short questions concerning participation in training courses, experience in using AI, knowledge or application of UDL, and perceived level of preparedness. The third includes four open–ended questions on the definition of AI applied to teaching, opportunities in heterogeneous classrooms, the conditions through which AI may foster or hinder inclusion, and ethical, educational, and social risks. The fourth includes three open–ended questions concerning the competences to be developed, the forms of support considered useful, and any additional observations.

Socio–professional information and short questions were used exclusively to describe and contextualize the group; the core of the investigation consists of the open–ended responses. The dataset analysed contains no direct identifiers. Each participant was assigned a progressive code;

original formulations were preserved, while empty cells and technical labels devoid of semantic content were excluded.

The corpus was subjected to qualitative categorical analysis with an inductive orientation. After a full reading by question and by participant, the texts were segmented into units of meaning, to which descriptive codes close to the respondents' own vocabulary were assigned. The codes were compared, grouped, and traced back to categories constructed a posteriori, verifying their internal coherence, conceptual distinction, and capacity to represent minority or dissonant positions as well. In the interpretive phase, the categories were placed in dialogue with the literature on Generative AI, UDL, teachers' digital competence, inclusion, and formative assessment. Convenience sampling, territorial circumscription, the self–reported nature of the data, the use of an ad hoc instrument, and the absence of independent double coding delimit the transferability of the results, which therefore have an exploratory value and are generative of hypotheses.

Results and discussion

The short questions outline the background against which the qualitative corpus should be read. Eighteen teachers report having already used AI tools, even occasionally, while 15 do not report any instructional use. Eight participants state that they are familiar with UDL. Perceived preparedness is distributed as follows: 15 negative responses, 6 partial or insufficient responses, 10 positive or moderately positive responses, and 2 ambiguous or dissonant responses. These counts have a descriptive, not inferential, function.

The first category, AI as operational and instructional support, gathers definitions centred on help, means, resource, and tool. Teachers refer to information search, lesson preparation, the construction of materials, outlines and grids, test design, and reduction of working time. In numerous responses, the value of AI remains anchored to operational efficiency; references to a transformation of learning goals or processes are less frequent. The corpus also preserves markedly distrustful positions, in which AI is described as lacking creativity, conditioning, or incompatible with the development of autonomous thinking. This ambivalence is consistent with the literature, which places the opportunities of co–design alongside the need for critical selection and teacher mediation (Evmenova et al., 2024).

The second category, inclusion as a conditioned potential, includes the personalization of learning pathways, adaptation and simplification of materials, differentiation of activities, alternative explanations, support for students with special educational needs or language difficulties, and timely feedback. AI is considered inclusive when it integrates the teacher's work, is contextualized, and remains accessible to all (Ferrari & Ellerani, 2025); conversely, it becomes a possible barrier when it replaces educational action, generates inaccurate content, or accentuates the digital divide. The gap between declared knowledge of UDL, limited to eight participants, and the more widespread presence of concepts related to personalization and barrier reduction suggests an inclusive sensitivity that is not yet fully organized within a shared design language. This interpretation is consistent with the UDL Guidelines 3.0 and finds a parallel in the evidence on AI–based assistive technologies, although the latter cannot be automatically extended to GenAI or to the Italian context (CAST, 2024; Anwar et al., 2026).

The third category concerns risks to cognitive autonomy, knowledge reliability, and the educational relationship. The responses refer to dependence, delegation of the task, plagiarism, reduced effort, weakening of critical thinking and creativity, privacy, bias, disinformation, the digital divide, and the possible thinning of the human relationship. One teacher summarizes the risk as the habit of "not thinking with one's own head"; another emphasizes that AI can facilitate inclusion if it integrates educational work, but may hinder it if it becomes a substitute for it. These concerns intersect with the empirical data reported by Agrati and Beri (2025) regarding teachers' worries about the loss of cognitive abilities, students' technological dependence, and the potential generation of errors by algorithmic systems. These concerns also intersect with the ethical and human–centred dimensions indicated by the European Commission and UNESCO (European Commission, 2022; UNESCO, 2024).

Feedback is also perceived as promising only when it remains embedded in explicit objectives, transparent criteria, and human mediation, in line with Huesca et al. (2025).
The fourth category, fragility of perceived preparedness and demand for situated training, shows that needs are not exhausted by the technical use of tools. Participants call for competences to understand how systems work, select reliable applications, evaluate sources and outputs, recognize errors and bias, address ethical and regulatory aspects, design discipline–specific activities, personalize materials, and assess assignments produced with the support of AI. The most frequently mentioned forms of support are practical courses, workshops, discipline–specific examples, peer exchange, guidelines, reliable tools, and accompanying figures. This result converges with Granström and Oppi (2025), who identify preparedness, perceived usefulness, and organizational support as relevant conditions for AI adoption.
Taken together, the categories make it possible to formulate, in heuristic and preliminary terms, Inclusive Pedagogical Readiness for AI (IPR–AI): a professional capacity that integrates critical understanding of systems, design oriented toward barrier reduction, protection of agency and human mediation, and reflective observation of effects on practices and learning. The data do not measure or validate this construct; rather, they support an initial interpretive articulation to be subjected to subsequent verification.

Conclusions

The preliminary findings respond to the research questions by showing plural professional representations: AI is interpreted as an operational support, a resource for personalization and accessibility, but also as a possible factor of dependence, inequality, and cognitive impoverishment. Its inclusive potential does not emerge as an intrinsic quality of the technology; rather, it depends on the teacher's ability to recognize barriers, verify outputs, differentiate proposals, and maintain educational and assessment responsibility (CAST, 2024; Evmenova et al., 2024). The gap between instrumental familiarity and pedagogical competence constitutes the most relevant finding. The needs expressed simultaneously concern technical literacy, critical evaluation, inclusive design, feedback management, and ethical and regulatory issues. This points to the need for professional development pathways that are workshop–based, discipline–specific, and collegial, supported by guidelines and organizational accompaniment (UNESCO, 2024; Granström & Oppi, 2025). The original contribution of the study lies in the preliminary proposal of IPR–AI as an interpretive lens for teachers' professionalism in the algorithmic environment. It does not constitute a scale or a validated measure of readiness, but rather a heuristic model constructed through the encounter between empirical categories and the literature, useful for guiding subsequent research and training devices.The findings must be read in light of convenience sampling, the predominantly regional scope, the self–reported nature of the responses, the use of an ad hoc instrument, and the absence of independent double coding. Future developments should expand the corpus, involve different geographical areas and school levels, consolidate the instrument, introduce procedures of intersubjective analysis, and complement declared representations with the observation of practices, the analysis of artefacts, and students' voices. From this perspective, the operational power of AI can become authentically educational when it remains guided by pedagogical judgment, professional responsibility, and care for differences.

Keywords: AI literacy; Inclusive Pedagogical Readiness; Instructional accessibility; Professional development

Digital Competence, Generative AI and Large–Scale Assessment: A Comparative Analysis of DigComp 3.0 and ICILS 2023

Marco Giganti

Introduction

Digital competence has become a central construct in educational research, policy and curriculum development. The fast diffusion of digital technologies, platform–based learning environments and generative artificial intelligence (GenAI) has transformed digital competence from a set of technical skills into a complex educational construct that includes cognitive, communicative, ethical, critical and reflective dimensions. In school contexts, students do not only need to access and use digital tools; they also need to interpret digital information, evaluate algorithmically mediated content, create digital artefacts, protect their well–being, understand data–driven environments and act responsibly in increasingly hybrid socio–technical contexts (Selwyn, 2022; Williamson et al., 2020, Rivoltella & Rossi, 2024).

The spread of GenAI intensifies this challenge. Large language models and AI–powered educational systems are progressively entering teaching, learning and assessment practices, offering opportunities for personalized learning, automated feedback, adaptive assessment and data–informed decision–making. At the same time, they raise significant questions concerning authorship, epistemic agency, transparency, bias, privacy, reliability of information and the educational meaning of assessment (UNESCO, 2023; European Commission, 2022). These issues require schools to move beyond a narrow operational view of digital competence and to address the broader conditions under which learners interact with digital and AI systems.

Against this background, the relationship between competence frameworks and assessment frameworks becomes particularly important. Frameworks do not simply describe educational constructs; they shape how educational systems define what is worth teaching, learning and assessing. At the European and international levels, two influential frameworks currently play a major role in this field. The Digital Competence Framework for Citizens, DigComp 3.0, developed by the European Commission's Joint Research Centre, provides a normative and curriculum–oriented model of digital competence (Cosgrove & Cachia, 2025). It describes the knowledge, skills and attitudes required for participation in society, work, learning and everyday life, and it explicitly incorporates recent technological developments, including AI, digital well–being, cybersecurity, sustainability, online identity and responsible participation. The International Computer and Information Literacy Study, ICILS 2023 (the latest available, the next is 2028), developed by the International Association for the Evaluation of Educational Achievement, offers a measurement–oriented framework for assessing students' Computer and Information Literacy (CIL) and Computational Thinking (CT) at Grade 8 through standardized, psychometrically robust instruments (Fraillon et al., 2024).

Although both frameworks address digital competence, they differ in purpose, scope and epistemic orientation. DigComp 3.0 articulates an aspirational and comprehensive view of what citizens should progressively develop across the lifespan. ICILS 2023 operationalizes selected dimensions of digital competence through tasks that can be assessed reliably and compared across educational systems. This distinction raises a crucial research problem: to what extent do normative competence frameworks and large–scale assessment frameworks align, and what are the implications of their convergences and divergences for teaching, curriculum design and assessment in the age of GenAI?

This contribution addresses this question through a systematic comparative analysis of DigComp 3.0 and ICILS 2023. The study examines the conceptual overlaps, structural differences and areas of misalignment between the two frameworks. It pays particular attention to the extent to which large–scale assessment frameworks can capture emerging dimensions of digital competence, including AI literacy, digital citizenship, well–being and meta–competence for lifelong learning. The paper

therefore contributes to the educational research by critically discussing how digital competence can be conceptualized, taught and assessed when GenAI and data–driven technologies are redefining educational practices.

Data used

The study is based on documentary data drawn from two official framework documents. The first source is DigComp 3.0 (Cosgrove & Cachia, 2025), the fifth edition of the European Digital Competence Framework for Citizens. This framework was selected because it represents one of the most influential normative references for digital competence in European education and policy. It structures digital competence into five areas and twenty–one competences and provides descriptors, learning outcomes and examples that can support curriculum design, professional development and policy implementation.

The second source is the ICILS 2023 Assessment Framework (Fraillon et al., 2024). ICILS was selected because it is one of the most relevant international large–scale assessments focused on students' digital competence. It assesses Computer and Information Literacy and Computational Thinking and provides a structured model for translating digital competence into measurable constructs, assessment tasks and comparable indicators. The framework is particularly relevant for this study because it shows how educational systems operationalize selected aspects of digital competence within the constraints of standardized assessment.

The units of analysis were the twenty–one DigComp 3.0 competences and the strands and aspects of ICILS 2023 CIL and CT. The analysis did not use student microdata or country–level ICILS results. Instead, it focused on the conceptual and operational architecture of the two frameworks. This choice is coherent with the aim of the study, which is not to estimate students' performance but to examine how digital competence is defined, delimited and made assessable across two major framework traditions.

Additional scholarly and policy literature on digital competence, educational assessment, AI in education and datafication was used to contextualize the comparison. This literature supported the interpretation of the frameworks within broader debates on digital transformation, GenAI, educational governance and the relationship between curriculum expectations and assessment practices.

Methodology

The study adopts a qualitative comparative methodology combining document analysis, competence–level mapping and thematic synthesis. This approach was chosen because DigComp 3.0 and ICILS 2023 are not directly equivalent instruments (Bowen, 2009). They differ in function, structure and intended use: the former is a normative framework for citizens' digital competence, while the latter is an assessment framework designed to support international measurement and comparison. A qualitative comparative approach makes it possible to examine these differences without reducing one framework to the logic of the other (Gale et al., 2013).

The first phase consisted of an in–depth document analysis of the two official frameworks. The analysis focused on four dimensions: the definition of digital competence; the architecture of the framework; the intended educational or assessment function; and the role attributed to teaching, learning and evaluation. Particular attention was paid to how each framework defines the boundaries of digital competence and how it integrates technical, cognitive, ethical, communicative and social dimensions.

The second phase involved the construction of a systematic crosswalk (Thomas & Harden, 2008). Each of the twenty–one DigComp 3.0 competences was mapped against the strands and aspects of ICILS 2023 Computer and Information Literacy and Computational Thinking. The mapping was guided by three analytical criteria. The first criterion was semantic equivalence, understood as the degree of conceptual alignment between the definitions and descriptors used in the two frameworks. The second criterion was structural equivalence, referring to the position occupied by a competence within the overall architecture of each framework. The third criterion was functional equivalence,

defined as the extent to which a competence can be operationalized through observable tasks or assessment indicators.

On the basis of these criteria, each relationship was coded as strong, partial or no correspondence. Strong correspondence was assigned when a DigComp competence had a clear conceptual and functional equivalent in ICILS. Partial correspondence was assigned when ICILS covered only some aspects of the DigComp competence or translated it into a narrower assessment construct. No correspondence was assigned when a competence was absent from the ICILS cognitive framework or appeared only indirectly in contextual materials rather than as an assessed construct.

The third phase consisted of thematic synthesis. The crosswalk results were reorganized to identify broader patterns of convergence and divergence. This step moved beyond individual competences and focused on the overall conception of digital competence embedded in the two frameworks. The synthesis allowed the identification of a shared measurable core, areas of partial translation and dimensions that remain difficult to assess through current large–scale assessment models.

Results and discussion

The analysis shows a strong convergence between DigComp 3.0 and ICILS 2023 in some core domains of digital competence. The clearest alignment concerns information and data literacy. Both frameworks emphasize the ability to search for, access, manage, evaluate and use digital information. In ICILS, this domain is central to Computer and Information Literacy and is operationalized through tasks requiring students to locate information, assess its relevance and credibility, organize digital resources and produce information–based outputs. This area therefore represents a consolidated and assessable component of digital competence.

A second area of convergence concerns digital content creation. DigComp 3.0 includes the ability to develop, integrate, re–elaborate and share digital content, while ICILS assesses students' capacity to produce digital information products in structured task environments. Although ICILS necessarily narrows the scope of content creation to tasks compatible with standardized assessment, both frameworks recognize that digital competence involves active production and communication, not only access or retrieval.

A third area of strong alignment concerns computational thinking. DigComp 3.0 includes problem–solving and the capacity to interact critically with digital environments, while ICILS 2023 explicitly assesses Computational Thinking as a distinct domain. This convergence is significant because it indicates that computational approaches to problem–solving have become part of the stable core of digital competence in both normative and evaluative frameworks.

At the same time, the analysis reveals important divergences. Several dimensions that are central in DigComp 3.0 are only partially represented or largely absent from the ICILS cognitive assessment framework (Cosgrove & Cachia, 2025; Fraillon et al., 2024). These include digital citizenship, online identity management, digital well–being, environmental sustainability, AI literacy and meta–competence for lifelong learning. These dimensions are highly relevant for education, especially in the context of GenAI, but they pose considerable challenges for standardized assessment. They involve ethical judgment, reflective awareness, contextual decision–making and long–term dispositions that are difficult to capture through short, controlled assessment tasks.

AI literacy represents a particularly important case. DigComp 3.0 integrates AI transversally as part of what citizens need to understand and manage in contemporary digital environments. This includes awareness of algorithmic systems, critical engagement with automated outputs, responsible interaction with AI tools and understanding of risks related to bias, opacity and misinformation. ICILS 2023, by contrast and waiting for the new framework release, primarily assesses CIL and CT through tasks that measure students' ability to use information and computational procedures. While some ICILS tasks may indirectly involve skills relevant to AI–mediated environments, the framework does not fully capture the specific forms of epistemic agency required when learners interact with GenAI systems.

This divergence does not indicate a weakness of ICILS. Rather, it reflects the specific function of large–scale assessment and a difference in the time of frameworks release. International assessments must prioritize comparability, reliability and psychometric robustness. As a result, they tend to focus on constructs that can be operationalized through observable tasks and scored consistently across countries. Normative frameworks such as DigComp 3.0 serve a different purpose: they articulate a broader educational horizon and identify competences that educational systems should promote, even when those competences cannot yet be fully measured through standardized instruments (Grek, 2020; Williamson et al., 2020).

The findings therefore suggest that digital competence should not be reduced to what large–scale assessments can currently measure. Assessment frameworks provide essential empirical evidence, but they necessarily select and simplify. Curriculum frameworks, by contrast, can incorporate emerging educational priorities, including AI literacy, digital well–being and responsible participation. The two types of frameworks should therefore be read as complementary rather than competing. DigComp 3.0 helps define the broader educational aims of digital competence, while ICILS 2023 provides robust evidence on selected measurable components.

This complementarity has implications for teaching and assessment. Teachers need frameworks that help them design learning experiences addressing both technical–cognitive skills and broader ethical, communicative and reflective dimensions. At the same time, educational systems need assessment models that can provide valid evidence without narrowing the curriculum to what is easily measurable. In the age of GenAI, this balance becomes even more important. Students must learn to use AI tools critically and purposefully, but they must also preserve their epistemic agency, understand the limits of automated outputs and engage responsibly with digital environments.

The analysis also points to the need for more diversified assessment strategies. Large–scale assessments remain essential for monitoring system–level patterns and inequalities, but they should be complemented by classroom–based assessment, performance tasks, portfolios, scenario–based assessments and reflective activities. These approaches may be better suited to capturing complex dimensions such as AI literacy, digital citizenship and ethical decision–making. The challenge is not to replace psychometrically robust assessment, but to integrate it within a broader ecology of evidence.

Conclusions

The comparative analysis of DigComp 3.0 and ICILS 2023 highlights both convergence and tension between normative and assessment–oriented approaches to digital competence. The findings indicate that information literacy, digital content creation and computational thinking constitute a relatively stable core of digital competence. These dimensions can be coherently addressed through curriculum–oriented frameworks and large–scale assessment instruments.

However, the analysis also shows that several dimensions considered educationally crucial in contemporary digital societies remain only partially represented in current large–scale assessment frameworks. AI literacy, digital citizenship, well–being, sustainability, online identity management and meta–competence for lifelong learning are central to DigComp 3.0 but remain difficult to translate into standardized assessment tasks. This gap is especially relevant in the context of GenAI, where students need not only operational skills but also critical judgment, ethical awareness and the capacity to negotiate meaning with algorithmic systems.

The study argues that these divergences should be interpreted as productive tensions rather than as inconsistencies. Normative frameworks and assessment frameworks answer different questions. DigComp 3.0 asks what citizens should know and be able to do in order to participate actively and responsibly in digital societies. ICILS 2023 asks what students can demonstrate under standardized assessment conditions and how educational systems compare with respect to selected digital competences. Both perspectives are necessary.

For research, the study contributes a structured crosswalk that can support future analyses of digital competence frameworks and assessment models. For policy, it highlights the need to align

curriculum expectations, teacher professional development and assessment practices without reducing digital competence to its most measurable components. For teaching, it suggests that educators should use large–scale assessment evidence as one source of information while also designing activities that address the broader reflective, ethical and AI–related dimensions of digital competence. In conclusion, the contribution supports a dialogical approach to digital competence research. Such an approach recognizes the value of empirical evidence from large–scale assessments while preserving the broader educational aims expressed by normative frameworks. This perspective can help schools respond to the challenges posed by GenAI and digital transformation, promoting forms of teaching and assessment that support students' critical participation, responsible agency and lifelong learning.

Keywords: Digital competence; Generative Artificial Intelligence; Large–scale assessment; Learning assessment

Classifying Items on the basis of Difficulty and Time–Intensity through Tree–Based Methods: Evidence from the INVALSI Digital Competence Assessment

Tommaso Giorgi – Luca Bungaro – Mariagiulia Matteucci – Stefania Mignani

Introduction

In educational assessment, item properties are usually evaluated through pre–testing procedures aimed at estimating psychometric measures such as item difficulty. However, these procedures can be time–consuming, costly, or unfeasible, especially when new items must be produced rapidly or when items cannot be administered to sufficiently large samples. For this reason, recent research has investigated whether observable item characteristics, including textual and structural features, can be used to predict item parameters and support item development.

Matteucci, Mignani and Veldkamp (2012) first showed that item characteristics can be used to predict their respective parameters through multivariate regression trees, providing support for item calibration and adaptive testing when direct calibration is costly or based on limited samples. More recently, Štěpánek, Dlouhá and Martinková (2023) compared machine–learning methods for predicting item difficulty from textual features, showing that these approaches can compete with expert–based predictions and may be useful when pre–testing is limited or unavailable.

Extending these lines of research, the present study investigates the use of classification trees to identify homogeneous item profiles defined by item difficulty and time demand. Our previous work (Giorgi et al, 2026) on INVALSI digital competence items used multivariate regression trees to analyse continuous estimates of item difficulty and time intensity obtained from a Log–Normal Response Time Item Response Theory model (Van der Linden, 2007; Fox and Marianti, 2017). In contrast, the present contribution adopts a classification framework, in which items are assigned to classes according to their categorical levels of difficulty and time intensity. Specifically, we propose three separate univariate classification trees: one for difficulty classes, one for time–intensity classes, and an additional one based on a combined categorical profile. The aim is to identify groups of items that are homogeneous with respect to difficulty and/or time demand, while preserving the interpretability of tree–based procedures. The item characteristics selected by the trees may then be used as structured input for Large Language Models in automatic item generation, allowing new items to be based on features associated with specific psychometric and time demanding profiles to reduce validation time when extensive pre–testing is not feasible.

Data

The empirical application is based on data from the 2025 INVALSI Digital Competence Assessment, administered to 19,020 Grade 10 Italian students in a computer–based testing environment. The

assessment was designed considering the DigComp 2.2 framework and included items covering four areas of digital competence: information and data literacy, communication and collaboration, digital content creation, and safety.

The unit of analysis is the test item. The final dataset includes $K = 246$ multiple–choice items, each described by a set of covariates and by two outcomes: item difficulty and time intensity. These two quantities were first estimated as item–level parameters of a Log–Normal Item Response Theory model, using students' binary responses (correct/incorrect) and response–time data recorded during the survey. The resulting estimates of item difficulty and time intensity were then used as outcomes in the subsequent classification–tree analysis.

For the present analysis, item difficulty and time intensity are each discretized into three ordered classes using the 33rd and 66th percentiles of their empirical distributions as cut–off points. This produces two categorical item–level outcomes: Y_k^D, denoting the difficulty class of item k (Easy, Medium, Advanced), and Y_k^T, denoting its time–demand class (Fast, Moderate, Slow). These two outcomes are first analysed separately through two univariate classification trees. In addition, their cross–classification (Y_k^{DT}) can be used to define a combined item profile: easy–fast, easy–moderate, easy–slow, medium–fast, medium–moderate, medium–slow, advanced–fast, advanced–moderate and advanced–slow. When the combined profile is considered, a third univariate classification tree is estimated using the 3×3 difficulty–time–demand profile as the categorical response variable.

	Fast	Moderate	Slow	Total
Easy	50	23	9	82
Medium	15	33	34	82
Advanced	17	26	39	82
Total	82	82	82	246

Table 1: Cross–classification of categorical difficulty and categorical time–intensity

The resulting cross–classification of difficulty and time demand provides the empirical basis for identifying joint item profiles. Rather than focusing only on one response variable, the analysis investigates which item characteristics are associated with homogeneous profiles of both difficulty and time intensity.

The item–level predictors include textual, structural, content–related and framework–related features. Textual variables describe the wording of the item stem and answer options, while structural variables capture the presence of pictures, technical terms, and scenario–based content. Additional indicators identify references to digital elements, such as artificial intelligence, search engines and applications. Finally, framework–related variables describe the DigComp area and whether the item mainly targets knowledge or skill components.

Methodology

As a first step, item difficulty and time intensity were jointly estimated for the $K = 246$ test items using a Log–Normal Item Response Theory model. This modelling framework, originally proposed by van der Linden, allows response accuracy and response times to be analysed jointly, while accounting for the association between students' latent ability and working speed.

Let $i = 1, \dots, N$ index students and $k = 1, \dots, K$ index items. For each student–item pair, the model considers both the binary response variable Y_{ik}, coded 1 for a correct response and 0 otherwise, and the response time T_{ik}. At the person level, the model assumes a latent ability parameter θ_i and a latent working–speed parameter ζ_i. At the item level, it estimates parameters describing both the psychometric and temporal characteristics of the items.

Response accuracy was modeled using a probit Item Response Theory (IRT) model:

$$P(Y_{ik} = 1|\theta_i, b_k) = \Phi(\theta_i - b_k)$$

where b_k denotes the difficulty parameter of item k, while Φ is the normal cumulative distribution function. Higher values of b_k indicate more difficult items that is, items requiring a higher level of latent ability to be answered correctly.
Response times are modeled through a log–normal model:

$$\log(T_{ik}) = \lambda_k - \phi_k \zeta_i + \epsilon_{ik}, \qquad \epsilon_{ik} \sim N(0, \sigma^2_{\epsilon_k})$$

where λ_k is the item time–intensity parameter, ϕ_k is the time–discrimination parameter, and ζ_i represents the working speed of student i. Larger values of λ_k identify more time–demanding items, as they indicate a higher expected response time after accounting for individual differences in working speed.
The model assumes conditional independence between response accuracy and response time given the latent person parameters, while the association between ability and speed is modelled at the population level. Similarly, the item parameters are allowed to be associated at the item level, making it possible to study the relationship between item difficulty, time intensity and time discrimination. Parameters were jointly estimated in a Bayesian framework using Gibbs sampling, as implemented in the LNIRT modelling approach (Fox, Klotzke and Simsek 2023).
The posterior estimates of b_k and λ_k were then used as item–level outcomes in the subsequent classification–tree analysis. Specifically, item difficulty and time intensity were discretized into ordered classes and analysed to identify item characteristics associated with homogeneous psychometric and time demanding profiles.
Tree–based methods (Breiman et al, 2012; Hastie, Tibshirani and Friedman, 2009) recursively partition the predictor space into a set of non–overlapping regions, with the aim of creating terminal nodes that are as homogeneous as possible with respect to the response variable. For each outcome (difficulty alone, time intensity alone and the joint output), the tree algorithm searches over the available item–level predictors and selects the split that maximally reduces node impurity. Standard impurity measures for classification trees include the Gini index and cross–entropy, both of which favour splits that produce child nodes with a clearer class composition.
The terminal nodes of the difficulty tree identify groups of items with similar psychometric demand, whereas the terminal nodes of the time–demand tree identify groups of items with similar response–time requirements. The third additional analysis allows us to explore whether specific item characteristics are associated with joint profiles, such as items that are easy but time–consuming, or items that are both difficult and time–demanding.

Results and discussion

The three different classification trees highlight the central role of textual and structural item features in distinguishing difficulty, time–demand, and joint item profiles.

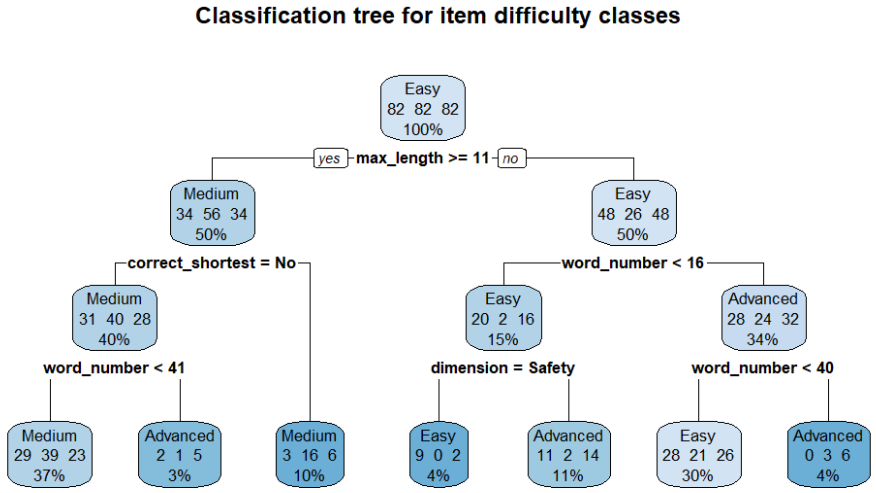


Figure 1: Classification tree of the categorical item difficulty

In the difficulty tree, the maximum length of the answer options is the first splitting variable, followed by features such as stem length and whether the correct answer is the shortest option. This suggests that the structure of the response alternatives and the amount of text to be processed are associated with different levels of psychometric demand.

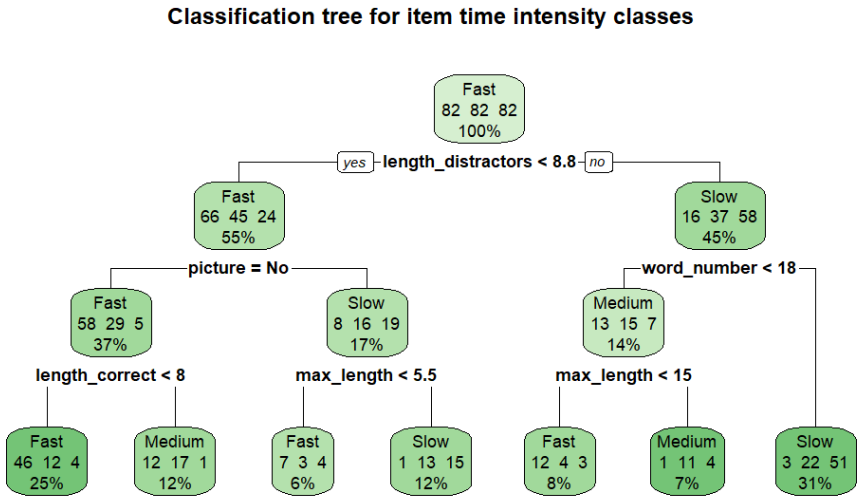


Figure 2: Classification tree of the categorical time–intensity.

The time–intensity tree shows a partly different pattern. The first split is based on distractor length, while subsequent splits involve the presence of pictures, correct–answer length, and stem length. Items with shorter distractors and shorter correct answers tend to be classified as fast, whereas longer textual components are more often associated with slow items. This supports the interpretation of time intensity as reflecting reading load and information–processing effort.

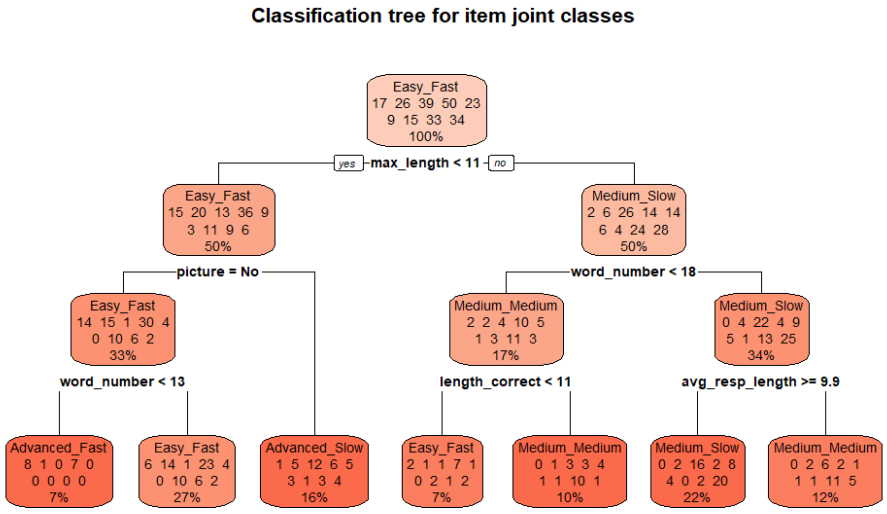


Figure 3: Classification tree of the joint categorical profiles.

Finally, the joint classification tree combines the two dimensions into a single profile. It confirms the relevance of answer–option length and stem length, while also showing that some item profiles cannot be fully understood from difficulty or time intensity alone. For example, items may be easy but slow, or more difficult but relatively fast. This joint perspective is useful for test design because it helps identify item characteristics associated with specific psychometric and temporal profiles, supporting the development of items that better balance difficulty and time burden.

Conclusions

This study shows how univariate classification trees can be used to identify item characteristics associated with homogeneous profiles of difficulty and time–intensity in the INVALSI Digital Competence Assessment. The results suggest that textual and structural features, such as answer–option length, stem length, distractor length, and the presence of pictures, play an important role in distinguishing both psychometric and time demanding item profiles. Moreover, the joint classification of difficulty and time intensity provides additional information for test design, highlighting items that may be easy but time–consuming, difficult but fast, or other combination.

From an applied perspective, these results may support item development by providing interpretable rules on the features associated with specific item profiles. Such information can also be used as structured input for automatic item generation, helping to produce new items with more predictable difficulty and time–demand characteristics. In this perspective, a relevant objective is to ensure adequate coverage of all difficulty–time–demand combinations, so that LLM–based item generation can be guided towards the full range of possible item profiles rather than only the most frequent ones.

Future work will extend this approach by developing and evaluating multivariate classification trees for categorical item outcomes. A key methodological issue concerns the definition of an appropriate impurity measure able to account for the joint distribution of difficulty and time–intensity classes, rather than treating them as separate outcomes or collapsing them into a single combined category. Further methodological work may also investigate tree–growing or post–processing strategies aimed at representing each relevant difficulty–time–demand combination in the terminal nodes, when supported by the data.

Fundings

This work was partially supported by the Italian Ministry of University and Research (MUR) through the PRIN 2022 project “Latent Variable Models and Dimensionality Reduction Methods for Complex

Data" (Project No. 20224CRB9E, CUP C53C24000730006, Grant Assignment Decree No. 1401, 18 September 2024).

Keywords: Digital Competences; Classification Trees; Item Difficulty; Response Time

SESSION 12. SUCCESSFUL SCHOOL–TO–UNIVERSITY TRANSITIONS: DETERMINANTS AND MEASURES TO CONTRAST UNIVERSITY DROP–OUT

ORGANIZERS: UNIVERSITY OF BRESCIA – UNIVERSITY OF BERGAMO
COORDINATORS: GIOVANNI ABBIATI – FEDERICA ORIGO
5TH NOVEMBER: 4.30 P.M. – 6.30 P.M. {AULA MAGNA – RESEARCH 13}

When Research Meets Teaching: Evidence from student level data

Annalisa Cristini – Federica Origo – Piera Bello – Costanza Marconi

Introduction

Universities are central institutions for both the production and the diffusion of new knowledge. Faculty members are simultaneously researchers, engaged in advancing the scientific frontier, and teachers, responsible for transmitting knowledge to students. While these two roles are formally complementary, the actual relationship between research and teaching remains an open em–pirical question. A growing body of work examines whether research activ–ity enhances or detracts from teaching quality (Rodr´ıguez and Rubio, 2016; Garc´ıa–Gallego et al., 2015). A key argument in favor of complementarity is that research–active instructors may introduce more recent ideas into their courses, thereby facilitating the diffusion of innovation within higher educa–tion (Biasi and Ma, 2025). At the same time, research productivity may also crowd out teaching effort (Braga et al., 2014) or interact non–linearly with pedagogical effectiveness (Art´es et al., 2017). Thus, understanding when and how research affects teaching remains essential for evaluating the role of the university as a vehicle of knowledge transmission. Prior work has linked stu–dent outcomes to various instructor characteristics, including subject–matter knowledge and formal qualifications (Gonz´alez–Regalado, 2025; Sancassani, 2023), cognitive skills and academic records (Bhai and Horoi, 2019; Hanushek et al., 2016), employment conditions such as tenure status (Figlio et al., 2015), and demographic match between students and instructors (de Gendre et al., 2024; Gershenson et al., 2022). A more recent strand of research explicitamines the dual role of university faculty as researchers and teachers by focusing on the effect of research productivity (Palali et al., 2018; Nur–tegin et al., 2020). Within this literature, findings remain mixed: research pro–ductivity may enhance teaching by bringing more current knowledge to the classroom, but it may also reduce effectiveness when high–impact research crowds out time and effort devoted to students (Coccorese et al., 2024).

We advance this literature by providing the first empirical evidence that the match between what instructors research and what they teach plays a central role in shaping student achievement in higher education. By shifting attention from how much faculty publish to what they research in relation to what they teach, we identify a previously overlooked channel through which new knowledge is transmitted to students.

Data

The empirical analysis relies on a novel matched student–teacher dataset for a mid–sized Italian university covering academic years 2015/16 to 2023/24, which combines administrative, textual, and survey data. Specifically, student–level administrative records are linked to the full textual content of all course syllabi taught at the university and to instructors' publication data from Sco–pus. We further complement these data with student–level survey information on post–graduation outcomes from AlmaLaurea.

Methodology

Research–Teaching Similarity

We use Natural Language Processing (NLP) to measure the semantic simi–larity between course syllabi and instructors' research outputs. Specifically, we transform each syllabus and publication abstract into a TF–IDF vector, which captures the importance of each term while downweighting very com–mon words that carry limited information.

We then compute cosine similarity between each course syllabus and each publication abstract, comparing their content based on the direction of the vectors. For each course, our similarity measure is defined as the maximum similarity between the syllabus and the instructor's publications.

Identification Strategy

We exploit within–student and within–teacher variation, an approach that mitigates the main sources of selection bias, including student sorting, in–structor teaching style, and time–invariant differences in research productiv–ity. We estimate the following linear model:

$Y_{ict} = \alpha + \beta \text{Similarity}jct + \mu_{ij} + \eta_t + \epsilon_{ict}$, (1) where Y_{ict} denotes student i performance, measured by the exam score

in course c taught by instructor j in year t and Similarityjct is the similarity

measure. μ_{ij} denote student by instructor fixed effects and we also include academic year fixed effects η_t to account for any time–specific change. Stan–dard errors are clustered at the student level.

We also analyze the effect of teaching–research similarity on longer–term academic and labor market outcomes by estimating an analogous specifica–tion at the student level. In this case, we aggregate the similarity measure across all courses taken by a given student.

Results

Our results show that greater alignment between an instructor's research agenda and the content of the courses they teach is associated with higher student performance. Students earn higher grades when taught by instruc–tors whose research is closer to the course material, even after controlling for research productivity, teaching load, course characteristics, and student background. Moreover, the effect is stronger for courses at more advanced levels and for students with lower prior preparation and less family resources, suggesting that alignment facilitates the diffusion of recent knowledge and supports students who benefit most from structured guidance. These findings indicate that the transmission of new ideas into higher education is not au–tomatic but depends critically on whether instructors teach what they know best.

As for longer–term outcomes, our results indicate that greater teach–ing–research similarity significantly increases the likelihood of graduating on time, as well as both the final graduation score and the dissertation evalu–ation. We also find evidence that greater alignment between teaching and research not only improves academic performance and graduation outcomes, but also translates into economically meaningful gains in early labor market career.

Conclusion

Overall, our findings provide new evidence on the research–teaching nexus, suggesting that alignment between teaching and research is not merely a feature of academic organization, but a meaningful determinant of student achievement. These results have implications for course allocation policies and for broader debates on how universities can balance research activity and teaching effectiveness.

Keywords: Education; Syllabi; Text Analysis; Teaching Quality

Same Problem, Different Causes: Heterogeneity in the Predictors of First–Year University Dropout

Giovanni Abbiati – Federica Origo – Xiao Li – Federico Segato

Introduction

University dropout is a first–order concern for both equity and the efficient use of public spending on higher education (Aina et al., 2022). In Italy the problem is concentrated in the early stage of University career: roughly two–thirds of dropouts occur within the first two years (ANVUR, 2018). Because attrition is front–loaded, information observed by the end of the first semester is, in principle, sufficient to flag most at–risk students while there is still time to intervene. The classic determinants literature frames dropout as the outcome of academic and social integration shaped by background, prior schooling, and institutional context (Tinto, 1975; Spady, 1970; Clerici et al., 2015). A growing literature exploits administrative data and machine learning to build *early–warning systems* (EWS) that predict dropout from enrolment and first–term records. Two strands are directly relevant. The first develops multilevel machine–learning early–warning models on a single, technically homogeneous institution (Cannistrà et al., 2022). The second prioritises *prediction over causal interpretation* on national or pooled samples and reports strong aggregate predictive performance, with first–year credits, high–school grades, and family income as leading features (Delogu et al., 2024). In both cases the headline object is a *single* pooled risk model and its *aggregate* accuracy.

This paper asks a different question. We keep prediction in the background and ask whether the **predictors of dropout are homogeneous across degree courses**. The motivation is institutional: unlike a technical university, the Universities of Bergamo and Brescia enrol an extremely heterogeneous student population spanning humanities, economics, law, languages, and engineering. In such a setting, a pooled model implicitly assumes that the same early signal — a low TOLC score, an unmet remedial course, a professional–school diploma, a given income band — carries the same weight everywhere. Yet a growing body of evidence suggests that the drivers of dropout differ by field of study and by supply–side course characteristics (Vergolini and Vlach, 2024; Gitto et al., 2016), which is exactly what a pooled model cannot capture. We test that assumption directly.

Our central finding is a tension between prediction and interpretation. Predictive power, which rests on early–stage predictors, does vary across courses, but the differences are modest. What varies far more is how the individual predictors operate: the same signal carries a very different weight from one course to another. The two universities display the same broad pattern, with some differences remaining even between comparable courses. A pooled EWS therefore looks adequate on an aggregate AUC criterion while being miscalibrated at the level where interventions are actually designed and delivered — the degree course. The policy implication is that early–warning rules should be calibrated *per course*, not *per institution*.

This delivers three contributions. First, we document predictive parity alongside coefficient heterogeneity using interpretable models, rather than reporting a single pooled accuracy number. Second, we exploit two generalist universities to show that the heterogeneity is systematic — related to course content, cost, and duration — rather than idiosyncratic to one institution. Third, we provide a directly usable lesson for the design of EWS in non–selective, multi–field universities.

Data

We use student–level administrative data from the Universities of Bergamo (UNIBG) and Brescia (UNIBS), covering first–year students enrolled in three–year bachelor's programmes (plus single–cycle Law). Data cover the entry cohorts from the 2017–18 to the 2024–25 academic year. UNIBG contributes eight departments (Law; Engineering and Applied Sciences; Management, Information and Production Engineering; Letters, Philosophy and Communication; Foreign Languages; Management; Economics; Human and Social Sciences); UNIBS contributes its economics, law, and

three engineering departments. Together the two universities span the full range from engineering and the quantitative sciences to economics, law, and the humanities, which is precisely what makes course–level heterogeneity observable.[7]

Predictors are grouped into three blocks, all observed **no later than the end of the first semester** so that prediction is operationally feasible and leakage–safe:

- **Socio–demographic:** sex, age at matriculation, citizenship, residence (in the municipality of the course), upper–secondary diploma type, final secondary–school grade, ISEE band (with a separate category for non–submission), and scholarship status.
- **Admission:** the TOLC entrance–test score and the assignment of Additional Educational Requirements (OFA) for students below the TOLC threshold. Because TOLC versions, content, and maxima differ across courses, raw scores are not comparable; we standardise each score (and the secondary–school grade) as a percentage of its maximum to place all students on a common metric.
- **First–semester career:** credits earned (CFU), grade–point average of passed exams, and an indicator for having passed no exams.

Outcome. The dependent variable is first–year *late* dropout. We separate *early* dropouts — students who formally renounce before the end of the winter exam session, before first–semester performance is observed — and exclude them, since they fall outside the population an EWS can reach. The modelled outcome combines (a) failure to re–enrol or formal renunciation by 30 September of the following year and (b) re–enrolment with fewer than 6 CFU earned in the *second* semester. Component (b) captures students who remain formally enrolled but have effectively disengaged; basing it on second–semester credits keeps the outcome leakage–safe with respect to the first–semester predictors. Dropout alone serves as a robustness specification.

Methodology

Our approach is based on two complementary families of models, both aimed at characterising how the drivers of dropout vary across courses.

The workhorse is a logistic regression of late dropout on the full set of first–semester predictors, with university, year, and department fixed effects, which reproduces the standard pooled early–warning specification. To identify heterogeneity, we interact the key predictors (TOLC, OFA status, high–school type, ISEE, and scholarship) with the degree course, so that each predictor's effect is free to vary across courses; joint Wald tests on the interaction terms then determine which effects are genuinely heterogeneous, and average marginal effects by course, with the two universities side by side, show the direction and size of each driver. A multilevel logit with department random effects complements this: a random intercept gives the intraclass correlation, separating heterogeneity in baseline risk from heterogeneity in the slopes, and a random coefficient lets the effect of a predictor itself vary across courses.

The second family is machine learning, used for interpretation as much as for prediction. Random forests, gradient boosting, and a generalised multilevel tree recover nonlinear structure and interactions without imposing a functional form; their variable–importance measures and tree splits reveal which predictors are decisive and how that ranking shifts from one course to another. The same models also provide the predictive benchmark: comparing per–course accuracy (AUC and calibration on a held–out sample) shows whether heterogeneity in the *drivers* translates into heterogeneity in predictive *accuracy*. Robustness checks re–estimate the models separately by

[7] The analysis uses cohorts from 2017 through the complete 2024–25 academic year, together with first-semester performance for the 2025–26 cohort. The results in Section 4 are preliminary: they come from the current specification and will be developed further through additional robustness checks, alternative variable definitions, and feature engineering on the existing data, including the machine-learning analysis described below.

university and replace the outcome with dropout alone and with an earlier (31 January) information cutoff.

Results (preliminary)

Predictive accuracy is roughly uniform across courses.

With full first–semester information the pooled model predicts late dropout well. Splitting the sample 70/30, fitting the model on the training set, and computing the area under the ROC curve within each department on the held–out set yields out–of–sample values between 0.79 and 0.88, a spread of about nine points. Law and the humanities are the most predictable (AUC near 0.88) and the engineering and economics departments the least (AUC near 0.79), but no course falls outside this narrow band. A single pooled early–warning model therefore predicts about equally well in every course: there is no course in which dropout is markedly harder or easier to predict.

The drivers, however, are sharply heterogeneous.

We test whether each key predictor's effect on dropout varies across departments by interacting it with the department and jointly testing the interaction terms. Homogeneity is rejected for every academic and admission predictor: first–semester credits ($\chi^2_{12} = 290, p < 0.001$), OFA status ($\chi^2_{29} = 314,\ p < 0.001$), the standardised TOLC score ($\chi^2_{12} = 46,\ p < 0.001$), and the secondary–school leaving grade ($\chi^2_{12} = 29, p = 0.004$). By contrast, the protective effect of holding a scholarship does *not* vary significantly across departments ($\chi^2_9 = 9.9, p = 0.36$). The heterogeneity is concentrated precisely in the academic–performance and admission signals on which early–warning systems rely most, while the one purely financial instrument behaves uniformly.

The estimated interactions show that the heterogeneity is not merely a matter of magnitude. The marginal effect of first–semester credits on dropout is several times larger in economics and management (for example, in the Brescia economics department) than in the engineering departments, so a given credit shortfall is a much stronger warning sign in some courses than in others. The TOLC score is protective where the test content is aligned with the curriculum but flat, or even perversely signed, where the entrance test bears little relation to the course, so its value as a screening signal is a function of test–content alignment rather than a universal constant. A pooled model that performs well on aggregate accuracy can therefore be miscalibrated in use, because it attributes the same meaning to the same signal everywhere.

A multilevel logit with department random effects makes this distinction explicit. With a random intercept only, the department–level intraclass correlation is small (ICC ≈ 0.02): once student characteristics are accounted for, departments differ little in their *baseline* dropout risk, which is exactly why a single pooled model predicts about equally well in each course. Allowing the TOLC effect to vary by department, however, yields a large and highly significant random–slope variance (likelihood–ratio $\chi^2_2 = 338,\ p < 0.001$). The courses thus differ far more in how strongly the predictors operate than in their overall level of risk: the heterogeneity that matters is in the slopes, not the intercepts.

Conclusions

In this paper we use detailed student–level administrative panel data for two generalist universities, and we show that the standard predictors of first–year dropout are not uniform across degree courses. Predictive accuracy is roughly constant, since a single pooled early–warning model predicts about equally well everywhere (department–level AUC between 0.79 and 0.88), but the *drivers* are sharply heterogeneous: the effects of first–semester credits, the assignment of remedial courses, the entrance–test score, and the secondary–school grade all reject homogeneity across courses, whereas the scholarship effect does not. A given credit shortfall or entrance–test result therefore carries a systematically different weight depending on the course. A pooled model that performs well on aggregate accuracy can thus be miscalibrated where it is used, because it attributes the same meaning to the same signal everywhere.

The practical implication is that early–warning systems in non–selective, multi–field universities should be calibrated and acted upon at the course level rather than the institutional level. The two–

university design shows the heterogeneity is systematic rather than institution–specific, strengthening the external validity of the estimates.
Two extensions follow naturally. First, machine–learning models (random forests, gradient boosting, and a generalised multilevel tree) can be used not only as a predictive benchmark but as a tool for further detection: their variable–importance measures and tree splits indicate which predictors are decisive within each course and can surface nonlinear, course–specific risk profiles that a linear interaction model does not capture. Second, a complementary design would identify the *causal* effect of the assignment of remedial courses by exploiting the TOLC threshold in a regression–discontinuity design (Aucejo and El Khoury, 2025), testing whether remedial requirements act as pedagogical support or as a discouragement signal.

Keywords: Digital self–efficacy, ILSAs, confirmatory factor analysis, measurement invariance

Starting Off on the Right Foot? Remedial Requirements and Academic Progression

Federico Segato – Xiao Li – Federica Origo – Giovanni Abbiati

Introduction

Italy faces a persistent challenge in higher education. Despite the expansion of access to university, the share of young adults holding a tertiary degree remains among the lowest in Europe. In 2024, only 31.6 percent of individuals aged 25–34 had completed tertiary education, compared with 44.1 percent in the EU (European Commission, 2025). This gap reflects limited school–to–university transitions, as well as difficulties in academic progression and completion. Among new entrants to bachelor's programmes, only 56 percent complete their degree within three years after the expected duration, compared with 70 percent on average across OECD countries (OECD, 2025). These figures suggest that increasing enrolment is not enough if students enter university with heterogeneous levels of preparation and struggle to progress through the first year, policies aimed at supporting early academic experience become central to improving completion.
A common institutional response aimed at reducing gaps in students' initial preparation is the assignment of remedial courses.[8] In the Italian context, remedial courses are intended to address knowledge gaps that universities consider essential for successfully undertaking a degree programme. Their objective is therefore to support students entering university with weaker preparation in core subjects. At the same time, the design of these requirements is highly decentralized: the relevant TOLC test, the subject–specific thresholds, and the organization of the remedial activities vary across degree programmes, departments, universities, and cohorts.
In this paper we study whether remedial courses help first–year university students progress through higher education or instead act as an early signal of academic difficulties. We focus on two medium–sized universities in Northern Italy and use administrative data covering the universe of first–year students enrolled between 2017 and 2025. The data combine detailed information on students' background (including family income, migration background, high–school track and final grade) with entry–test scores and complete university careers. This setting allows us to analyse both short–term outcomes, such as credits earned during the first semester, and longer–term outcomes, such as enrolment in the second year and dropout. Compared with existing evidence for Italy, which is mostly based on single–university settings and fuzzy assignment rules, our study exploits

[8] In the Italian university system, these requirements are generally referred to as Obblighi Formativi Aggiuntivi (OFA). They are assigned to first-year students who do not reach the minimum score in entry tests, such as the TOLC, in specific subject areas.

predetermined TOLC cutoffs that assign students to remedial requirements across a wide range of degree programmes, departments, cohorts, and universities. We further leverage changes over time in both the relevant cutoffs and the organization of remedial activities to examine how the design of remediation policies shapes students' academic trajectories.

The international literature on remedial and developmental education provides mixed evidence on whether remediation helps or hinders students' academic trajectories. Some studies find that remedial courses can support underprepared students by improving persistence, transfer, and completion, especially when remediation addresses specific academic gaps and is well targeted (Attewell et al., 2006; Bettinger and Long, 2009). Other studies, however, find limited or null effects on academic attainment and later labour–market outcomes, suggesting that remediation may impose additional requirements without substantially improving students' progression (Martorell and McFarlin, 2011; Scott–Clayton and Rodriguez, 2015). A third set of contributions shows that average effects may hide substantial heterogeneity. Remediation can be beneficial for students with very weak initial preparation, while students close to the placement threshold may experience delays, diversion from regular coursework, or costs from being assigned to a remedial track (Boatman and Long, 2018). Related work also emphasizes that the effectiveness of remediation depends on the accuracy of placement rules and on the design of remedial sequences, since misassignment and low completion of remedial requirements can weaken or even reverse the intended effects (Bailey et al., 2010; Scott–Clayton et al., 2014). Overall, this literature suggests that there is no general answer to whether remediation works: its effects depend on the institutional context, the targeting rule, and the way remedial programmes are implemented.

Studying remedial education in Italy is not simply a replication of the US evidence. In the Italian university system, remedial requirements are embedded in degree–specific admission rules and are often implemented with substantial variation across institutions, departments, and cohorts. A small but growing literature has examined their effects. De Paola and Scoppa (2014) use a fuzzy regression discontinuity design at the University of Calabria and provide early causal evidence on the effectiveness of remedial courses. Duchini (2017) studies a sharp regression discontinuity design in one Economics department in Northern Italy, while De Benedetto et al. (2022) extend the analysis of remedial education to longer–run academic and labour–market outcomes. Related work has also examined specific mechanisms, such as procrastination and academic behaviour (De Paola and Scoppa, 2015), or closely related remedial policies in secondary education (Battistin and Schizzerotto, 2019). Overall, however, the Italian evidence remains concentrated in a limited number of institutional settings, often involving a single university, department, cohort, or remedial design. Given that the international literature shows that remediation effects are highly context–dependent, this leaves open the question of how such policies operate in a broader and more heterogeneous university setting.

Data

We use student–level administrative data from the Universities of Bergamo (UNIBG) and Brescia (UNIBS), covering the universe of first–year students enrolled in three–year bachelor's programmes and single–cycle Law degrees over the period 2017–2025. UNIBG contributes information from eight departments: Law; Engineering and Applied Sciences; Management, Information and Production Engineering; Letters, Philosophy and Communication; Foreign Languages; Management; Economics; and Human and Social Sciences. UNIBS contributes information from Economics, Law, and three Engineering departments. Together, the two universities cover a broad range of fields, from engineering and quantitative disciplines to economics, law, languages, humanities, and social sciences. This institutional variation is central to our setting, as remedial requirements differ across universities, departments, degree programmes, and cohorts.[9]

[9] The analysis uses cohorts from 2017 through the complete 2024–25 academic year, together with first semester performance for the 2025–26 cohort.

The dataset combines three main sources of information. First, we observe students' socio–demographic characteristics, including sex, age at matriculation, citizenship, residence, upper–secondary school track, final secondary–school grade, ISEE income band (with an indicator for non–submission of ISEE) and scholarship status. Second, we observe admission–test information, including TOLC entrance–test scores and the assignment of remedial courses (Obblighi Formativi Aggiuntivi, OFA) to students scoring below programme–specific thresholds.[10] We reconstruct the remediation regime at the degree–programme–by–cohort level by linking each student to the relevant TOLC type, testing period, cutoff, subject area, and remedial requirement. Third, we follow students' university careers over time, observing credits earned, grade–point average among passed exams, and whether the student passed no exams, with information that can be disaggregated by semester and academic year.

Our main outcome is first–year dropout. When possible, we distinguish early formal withdrawals before the end of the winter exam session from later dropout, for which academic performance during the first year can be observed. We also analyse short–term academic progression, measured by credits earned in the first semester, and medium–term persistence, measured by enrolment in the second year.

Methodology

We estimate the causal effect of remedial requirements on students' academic outcomes by exploiting a sharp regression discontinuity design. The assignment rule is based on programme–year–specific TOLC thresholds: students scoring below the relevant cutoff are assigned an Additional Educational Requirement, while students scoring above the cutoff are not. Following the regression discontinuity framework introduced by Thistlethwaite and Campbell (1960) and widely used in applied economics (Imbens and Lemieux, 2008; Lee and Lemieux, 2010), our identifying assumption is that, in a sufficiently small neighbourhood around the threshold, students just below and just above the cutoff are comparable in both observed and unobserved determinants of academic progression. Under this assumption, any discontinuous change in outcomes at the cutoff can be interpreted as the causal effect of being assigned to remediation for students at the margin.

Our baseline specification estimates local linear regressions on either side of the cutoff. Let Y_{ic} denote the outcome of interest for student i, $R_{ic} = S_{ic} - \overline{S_c}$ the TOLC score of student i centered at the specific cutoff S_{ic} of their enrolled programme–cohort c, $D_i = R_{ic} < 0$ an indicator for assignment to remedial requirements and α_c a set of programme–cohort fixed effects to account for baseline differences across institutional margins. We estimate:

$$Y_{ic} = \alpha_c + \tau D_i + \beta_1 R_{ic} + \beta_2 (D_i \times R_{ic}) + \gamma' X_i + \varepsilon_{ic}$$

within an optimal bandwidth around the cutoff. The parameter τ captures the discontinuity in the outcome at the threshold and represents the causal effect of remedial assignment for students close to the cutoff. The main outcomes are first–year dropout, credits earned in the first semester, and enrolment in the second year. We estimate the model separately by relevant institutional margins and, when appropriate, pool observations across cutoffs after centring the forcing variable around each programme–specific threshold. The richness of our data also allows us to explore heterogeneity across fields of study, departments, universities, and student background characteristics. This is important because remedial requirements may not have a uniform effect. For example, mathematics remediation may operate differently in STEM fields than in humanities or social–science programmes.[11]

[10] Informations on the before-TOLC period for OFA are actually being collected (covering typology of tests and related thresholds).

[11] Some students complete their OFA in time while others never do, our Sharp RDD identifies the Intention-to-Treat (ITT) effect of being assigned a remedial requirement.

We will assess the validity of the design through standard RD diagnostic checks. First, we will test for discontinuities in predetermined covariates at the threshold, including gender, age, citizenship, family income, high–school track, and secondary–school grade. Second, we will examine whether the density of the forcing variable changes discontinuously at the cutoff, which would indicate potential manipulation or sorting around the threshold. Third, we will test the robustness of the results to alternative bandwidths, polynomial orders, kernels, and specifications with and without covariates, using robust bias–corrected inference and data–driven bandwidth selection following Calonico et al. (2014). Finally, since the universities changed some cutoff rules and the organization of remedial requirements over time, we will complement the baseline sharp RD with a difference–in–discontinuities analysis Grembi et al. (2016). This extension exploits both the discontinuity at the cutoff and changes in the assignment rule or remediation design across cohorts, allowing us to assess whether more stringent thresholds or different forms of remediation affect students' academic trajectories.

Expected Results

While the empirical analysis will determine the direction and magnitude of the estimated effects, the existing literature suggests that the impact of remedial requirements is theoretically ambiguous. If OFA operate primarily as academic support, assignment to remedial courses should help students close initial knowledge gaps, improve their readiness for university–level coursework, and increase academic progression. In this case, we would expect positive effects on credits earned and persistence, and negative effects on dropout. Conversely, if OFA act mainly as an early signal of academic non–readiness, or if they divert time and effort away from regular coursework, remedial assignment may reduce first–semester performance and increase the probability of early dropout. This distinction is central to our analysis, as it allows us to interpret the estimated discontinuities as evidence on whether remediation works primarily as a support mechanism or as a discouraging signal.

We will therefore examine both short–term and medium–term outcomes. Credits earned during the first semester capture immediate academic progression, but their interpretation may be ambiguous: a negative short–run effect could reflect either discouragement or the mechanical time cost of completing remedial activities. By contrast, enrolment in the second year and first–year dropout provide more direct evidence on whether remedial assignment improves persistence or instead accelerates exit from university. We will also distinguish, where possible, between very early formal withdrawals and later dropout, for which students' first–year academic performance can be observed. Finally, we will study heterogeneity across fields of study and student characteristics. First, we will compare STEM and non–STEM programmes, since remedial requirements in mathematics or quantitative skills may have different implications when these competencies are central to the degree programme. In STEM fields, remediation may be especially useful if it addresses binding skill gaps, but it may also provide a stronger signal of mismatch between students' preparation and programme requirements. Second, we will examine gender differences, particularly within STEM and quantitative fields, where negative academic signals may interact with beliefs about ability, confidence, and perceived fit (Cheryan et al. 2017). Third, we will explore heterogeneity by prior preparation and socio–economic background, using information on high–school track, secondary–school grade, family income, and migration background. This will allow us to assess whether OFA benefit students with weaker initial preparation or instead impose costs on students close to the assignment threshold.[12]

Keywords: Remedial education; Higher education; University dropout; Regression discontinuity design

[12] Our estimates speak to whether remediation helps students at the margin, not to its effect on the most weakly prepared students further below the cutoff.

School–to–University Transitions: A Hybrid Machine Learning study of Tracking and Inequality

Giulia Pietrangeli – Michele Marsili – Fiammetta Noccioli

Introduction

The transition from upper secondary education to university represents a critical juncture in students' academic and professional trajectories. Rather than operating as a linear progression based on intellectual merit, this transition functions as a complex social allocation mechanism where cognitive skills, socio–economic backgrounds, information asymmetries, and institutional constraints intersect. When capable students fail to enroll in tertiary education, they contribute to a systemic bottleneck that reproduces structural inequalities.

We distinguish between two broad categories of determinants: (i) academic competencies and cognitive skills, and (ii) socio–demographic and institutional background factors. While the former capture students' academic preparedness and performance, the latter encompass family background, cultural capital, and the institutional constraints—such as school tracking—that shape educational opportunities. This distinction allows us to quantify the relative contribution of each dimension to university enrollment outcomes. Contemporary literature refines this through the lenses of bounded rationality and systemic information barriers: disadvantaged students systematically overestimate financial costs and dropout risks while underestimating long–term returns, leading to preventive self–exclusion.

The rigid tracking mechanisms in the Italian school system further exacerbate these dynamics. Early enrollment in vocational and technical tracks serves dual, often conflicting purposes: while vocational tracks accelerate employment entry for vulnerable groups, they create a severe "scarring effect" that penalizes tertiary enrollment. The institutional signal provided by school track strongly attenuates the relationship between individual academic merit and subsequent enrollment.

This study addresses three key questions: (i) What is the relative importance of academic competencies versus socio–demographic and institutional factors in predicting university enroll–ment, and how does the predictive power of early–stage data compare to late–stage data? (ii) When does student risk become identifiable, and what is the optimal window for intervention?

(iii) What is the magnitude of the scarring effect imposed by different school tracks?

Data Used

We leveraged a large–scale longitudinal administrative dataset from the Italian National Institute for the Evaluation of the Education and Training System (INVALSI), representing a near–complete census of Italian upper secondary school graduates. The initial cohort comprised approximately 400,000 students, tracked from Grade 10 (2018–2019) through Grade 13. After

excluding students who dropped out or were held back, the final analytical sample included approximately 375,000 high school graduates fully eligible for university enrollment. Enrollment records were linked via encrypted identifiers from the National Student Registry and the Ministry of University and Research (MUR). The target variable was verified university enrollment in the subsequent academic year (2022–2023), with 59.4% enrolled and 40.6% not enrolled, necessitating techniques to address class imbalance in the predictive modeling phase.

The feature space was structured into three dimensions:

Academic Performance and Objective Competencies: Standardized Weighted Likelihood Estimate (WLE) scores in Mathematics and Italian at Grades 10 and 13, plus curricular grades (raw year–end grades on the 1–10 scale).

Socio–Demographic and Cultural Background : Economic, Social, and Cultural Status (ESCS)

index, gender, migration origin (native, first–generation, second–generation), parental education (pre–diploma, diploma, post–diploma), and geographic macro–area (North, Center, South, Is–lands).

Institutional Tracking and Structural Constraints: Five school tracks (Classical/Scientific

Lyceums, Other Lyceums, Technical–Economic, Technical–Technological, Vocational) and school management type (public vs private).
Missing data (paternal education: 16.5%, maternal: 13.9%, curricular grades: 8.0%) were handled via Multiple Imputation by Chained Equations (MICE) nested within the cross–validation pipeline to prevent data leakage.

Methodology

We employed a hybrid pipeline combining machine learning and causal inference. Hyperparameter tuning was performed via Optuna with Tree–structured Parzen Estimator (TPE) sampler and 3–fold cross–validation maximizing F1–score. The Synthetic Minority Over–sampling Technique (SMOTE) balanced the minority class strictly within cross–validation folds to prevent data leakage.
Machine learning–based early warning systems have demonstrated strong predictive performance in educational contexts, particularly using Random Forest with administrative data. Four architectures were evaluated: Random Forest, Logistic Regression, Multi–Layer Perceptron, and TabNet.The Random Forest classifier emerged as optimal (AUC = 0.8163), offering superior predictive performance, interpretability, and calibration (Expected Calibration Error, ECE = 0.044) compared to Logistic Regression, TabNet, and MLP architectures.
To isolate the causal effect of school tracking, we applied Propensity Score Matching (PSM) with a 1:1 nearest–neighbor algorithm without replacement, caliper of 0.2 standard deviations of the logit (0.3170), and stratified subsampling (50,000 treated, 50,000 control, seed 42) preserving geographic distribution. Covariates included WLE scores, ESCS, parental education, gender, and macro–area. The Average Treatment Effect (ATE) was estimated with bootstrap confidence intervals. Rosenbaum bounds assessed sensitivity to unobserved confounding, with balance diagnostics confirming post–match standardized mean differences below 0.1 for all covariates.

Results and Discussion

1.1. Predictive Performance and Temporal Window

The Random Forest achieved AUC = 0.8163 (95% CI: [0.8134, 0.8195]) with Expected Calibration Error (ECE) = 0.044, outperforming Logistic Regression (0.071), TabNet (0.050), and MLP (0.085). The RF model minimized false negatives (9,427 errors) compared to MLP (12,790), TabNet (10,368), and Logistic Regression (11,731). In the context of educational policy, the cost of failing to identify vulnerable students outweighs false positives.
The Grade 10 model achieved AUC = 0.797, while Grade 13 added only 0.019 AUC, suggesting early intervention is feasible. However, Grade 10 predictions partially capture tracking assigned at Grade 8–9, indicating truly preventive intervention should begin before track assign–ment—during lower secondary school—to address root causes rather than downstream manifestations.

1.2. Relative Contribution of Competence vs. Background Factors

We aggregated SHAP values by computing the mean absolute contribution of each feature and summing within theoretically defined groups. Socio–demographic and institutional background factors accounted for 47.4% of predictive output, dominating academic competencies (43.6%), with additional demographic features contributing 9.0%. This represents a core contribution: inequality in university access is driven not by raw academic merit, but by students' origins and institutional sorting at age 14. This imbalance reflects the cumulative disadvantage embedded in the Italian educational system: by age 14, students from lower socio–economic backgrounds are already channeled into tracks that constrain their future opportunities, a process that late–stage cognitive remediation alone cannot fully reverse.

1.3. The Scarring Effect of Institutional Tracking

To quantify the scarring effect associated with different school tracks beyond correlational pat–terns, we employed Propensity Score Matching (PSM) to isolate the causal effect of tracking on university enrollment.

PSM estimated the ATE of vocational/technical track at −0.2565 (95% CI: [−0.2631, −0.2507]). Disaggregation revealed a clear gradient:

Table 1 : Disaggregated Average Treatment Effects by Track Type

Track	ATE	95% CI
Technical-Economic	-0.181	[-0.190, -0.171]
Technical-Technological	-0.245	[-0.255, -0.237]
Vocational	-0.362	[-0.370, -0.354]

Vocational students faced more than double the scarring effect of their Technical–Economic peers. This monotonic pattern indicates the institutional funnel operates as a graduated sorting mechanism, implying that the intensity of targeted interventions must be calibrated to specific track types, with the most comprehensive resources directed toward Vocational students. Rosenbaum bounds showed $\Gamma > 4.0$, indicating high robustness to hidden bias—an unobserved confounder would need to increase odds of track assignment by over 400% to nullify the result. Longitudinal academic trajectories confirmed the bias: a lyceum student with consistently low math scores still enrolled at higher rates than a vocational student with consistently high math scores, demonstrating the institutional label strongly attenuated the relationship between academic merit and enrollment. Beyond tracking effects, the model also exhibited systematic disparities across demographic groups. At the 0.5 threshold, females showed predicted enrollment of 0.691 (actual 0.656), males 0.500 (actual 0.539), natives 0.621 (actual 0.614), second–generation immigrants 0.421 (actual 0.483), and first–generation immigrants 0.293 (actual 0.407). School track disparities were stark: Lyceum predicted 0.978 (actual 0.863), Other Lyceum 0.750 (actual 0.646), Tech–Eco 0.327 (actual 0.476), Tech–Tech 0.255 (actual 0.416), and Vocational 0.001 (actual 0.164). Crucially, these disparities reflected the model's faithful representation of existing structural inequalities rather than algorithmic amplification. The model's predictions tracked observed demographic gradients without systematic exaggeration: the vocational–lyceum gap in predictions (0.001 vs 0.978) reflected the actual disparity (0.164 vs 0.863).

However, the standard 0.5 threshold—which assumes equal class probabilities—is inappropriate for Vocational students, whose actual enrollment rate is only 16.4%. The model's underprediction at this threshold thus necessitates track–specific adjustments ($P \leq 0.35$) in the 3–Tier Early Warning System (EWS) framework. The intersectional nature of disadvantage is evident in the gender–migration interaction: first–generation immigrant males enrolled at 36.2%, compared to 57.4% for native males—a gap of 21.2 percentage points that exceeds both the gender gap and the migration gap individually, reflecting a compounding effect of multiple marginalized identities.

Threshold Analysis and Policy Implications

Translating predictive probabilities into actionable interventions requires examining how thresh–old choices affect the identification of at–risk populations. We evaluated a 3–Tier stratification: Tier 1 (Low Risk, $P \geq 0.65$) containing 42.4% of the cohort, Tier 2 (Moderate Risk, $0.35 < P < 0.65$) containing 31.9%, and Tier 3 (High Risk, $P \leq 0.35$) containing 25.7%. These thresholds are

illustrative; optimal cutoffs depend on resource availability and must be validated through pilot implementations.

Tier 3 exhibited 77.8% non–enrollment, ESCS –0.43, and 69.3% vocational students. Within Tier 3, a "Lost" sub–segment ($P < 0.2$) existed, composed of 69.3% vocational students—confirming these individuals are structurally excluded rather than indecisive.

Cognitive remediation proved limited: a +0.5 SD math improvement among Tier 3 students increased enrollment probability by only 2.67%, rescuing 12.9% (2,488 of 19,305) from the high–est risk tier. This suggests systemic structural constraints prevent marginalized students from translating academic recovery into institutional access.

A sensitivity sweep of the uncertainty boundaries revealed that expanding the interval (e.g., 0.3–0.7) disproportionately captures students from vocational tracks and lower ESCS back–grounds. For example, a narrow zone (0.45–0.55) captures only 10.5% of the cohort with just 0.6% from vocational tracks; expanding to 0.2–0.8 captures 62.7% of the cohort but increases the vocational proportion to 5.7%. This demonstrates that model uncertainty is not random noise but reflects structural marginality.

Territorial analysis confirmed the traditional North–South divide is largely mediated by tracking and socioeconomic factors, with only a small residual territorial effect remaining after controlling for observable characteristics.

A governance framework for EWS implementation addresses data privacy, territorial partner–ships, and consent mechanisms aligned with the General Data Protection Regulation (GDPR) and Italy's National Recovery and Resilience Plan (PNRR), including data–sharing agreements between INVALSI/MUR and universities, aggregated risk profiles for resource allocation, and partnerships with local social cooperatives.

Conclusions

This study demonstrated that tertiary educational disparities in Italy are overwhelmingly driven by socio–demographic and institutional factors rather than cognitive merit. The tracking system imposes a graduated scarring effect intensifying with vocational orientation, with Vocational students facing a –0.362 ATE—more than double that of Technical–Economic peers. Grade 10 data enables early warning with AUC 0.797, but truly preventive action requires intervention before track assignment (Grade 8–9).

The 3–Tier EWS framework provides a scalable, fairness–aware blueprint for university governance, integrating territorial partnerships with social cooperatives. While findings confirm established patterns, the contribution of this study lies in the quantification of these mechanisms and the ability to simulate policy scenarios—bridging predictive analytics with actionable, data–driven intervention.

Several limitations must be acknowledged: the current dataset lacks post–enrollment university–level variables; the scarring effect is context–specific to the rigid Italian tracking system; and implementation requires fairness–constrained threshold adjustments. Additionally, while PSM balances observed covariates, the assumption of unconfoundedness remains a limitation, as un–observed factors like intrinsic student motivation are not captured in administrative data. De–spite these limitations, the methodological framework is highly transferable to other institutional settings equipped with longitudinal administrative data.

Keywords: Educational transitions; School tracking; Machine learning; Early warning systems

DROP–IT: Dropout Prevention in Italian Tertiary Education

Marta Campagnoli – Marta Magnani – Chiara Masci – Silvia Salini

Introduction

In Italy, tertiary educational attainment among young adults remains comparatively low: only around one third of people aged 25–34 hold a tertiary qualification, a share below the European/OECD average. Increasing the number of graduates therefore represents a key priority for both the higher education system and public policy (OECD, 2024; Eurostat, 2026). At the same time, the Italian university system is facing the consequences of demographic decline. The expected reduction in the size of younger cohorts is likely to decrease the pool of potential university entrants in the coming years, making student recruitment, retention and completion increasingly important for universities (ISTAT, 2024; ANVUR, 2024).

Against this background, preventing dropout is not only a matter of individual student support, but also a strategic issue for the sustainability of the higher education system. As participation widens and student trajectories become more diversified, universities need to better understand when and why students slow down, move across programmes or institutions, or leave higher education altogether. Previous evidence on the Italian higher education system shows that dropout is particularly relevant in the early stages of university careers, making timely identification and intervention especially important (ANVUR, 2018; Masci, Cannistrà and Mussida, 2024).

In this context, the study of dropout dynamics and student mobility is particularly relevant. It allows us to distinguish between trajectories that are often conflated in administrative data: permanent exit from the tertiary education system, mobility across degree programmes and universities, and mobility between traditional face–to–face universities and distance–learning universities. It also provides a more precise understanding of the timing and factors associated with different academic outcomes, which is essential for developing effective dropout–prevention policies and targeted support measures.

The aim of this study is to analyse the factors associated with different academic career outcomes among university students, including graduation, dropout and mobility across programmes or institutions. Particular attention is devoted to the role of students' background at enrolment, including previous educational pathways, type of upper–secondary school, high–school achievement and socio–demographic characteristics, which are crucial dimensions in the transition from school to university.

To this end, we apply survival models for competing events, which are statistical modelling techniques suited to the analysis of longitudinal academic trajectories. Survival analysis is particularly appropriate because it models the time until an event occurs and can account for censored observations, while competing–risks models allow different outcomes, such as dropout, graduation, transfer or institutional mobility, to be treated as distinct events rather than as a single undifferentiated outcome (Kleinbaum and Klein, 2005; Fine and Gray, 1999; Putter, Fiocco and Geskus, 2007). Given the nested structure of the data, with students enrolled in different degree programmes and institutions, we also explore multilevel and frailty specifications to account for heterogeneity across degree programmes, universities and geographical areas. Previous research has shown that programme–level heterogeneity can play an important role in dropout dynamics and in the timing of student withdrawal (Masci, Cannistrà and Mussida, 2024)

Data

We use data from the National Registry of Students (Anagrafe Nazionale Studenti, hereafter ANS), made available through an agreement between the Milan Higher Education Observatory (MHEO) and the Italian Ministry of University and Research.

The ANS provides administrative longitudinal data on students enrolled in Italian higher education institutions. For each student career, the dataset includes information on students' socio–demographic and educational background, such as gender, age, area of residence, type of upper–secondary school attended and final high–school grade. It also includes detailed information on university careers, such as the university and degree programme of enrolment, degree level and field of study, type of institution, academic year of first enrolment, credits earned over time, changes in programme or institution, transfers, withdrawal, graduation and other career outcomes.

Access to a dataset with national coverage allows us to conduct a more complete and systematic analysis of dropout and student mobility across the Italian university system. This is a crucial advantage compared with studies based on a single institution, where leaving a university may be observed but cannot always be clearly distinguished from moving to another institution or programme. By following students across institutions and academic years, the ANS makes it possible to reconstruct academic trajectories over time and to provide a more precise definition of dropout, distinguishing between students who leave the tertiary education system and students who move across degree programmes, universities or types of institutions. This informative richness is particularly relevant for analysing mobility between face–to–face and distance–learning universities, as well as movements across programmes with different levels of selectivity.

Methodology

The work is structured into three main phases. First, we build and prepare the analytical dataset using the National Registry of Students. This phase involves reconstructing students' academic careers longitudinally, harmonising information across academic years and institutions, and defining the main outcomes of interest: dropout from the tertiary education system, graduation, mobility across degree programmes or universities, and mobility between face–to–face and online universities.

Second, we develop descriptive and graphical analyses of academic trajectories and mobility dynamics. This phase aims to provide an initial mapping of students' pathways, with particular attention to transitions across programmes, institutions and types of universities, as well as to the timing of dropout and other career events. Descriptive analyses will also explore how these trajectories differ according to students' background at enrolment, including upper–secondary school track, high–school performance, geographical origin and other socio–demographic characteristics.

Third, we apply statistical modelling techniques to analyse dropout and mobility dynamics while accounting for the longitudinal and nested structure of the data. In particular, we test and compare survival models for competing events, including cause–specific hazard models and subdistribution hazard models. We follow students from their first enrolment to the end of a defined observation window and model the occurrence and timing of different academic outcomes, taking into account individual characteristics measured at enrolment, such as gender, age, citizenship, geographical origin, type of upper–secondary school and final high–school grade.

We also explore multilevel and frailty specifications to account for heterogeneity across degree programmes, institutions and geographical areas. This approach allows us to identify profiles associated with different academic trajectories and to investigate how dropout and mobility patterns vary across educational contexts.

Expected Results and Discussion

We expect to produce a detailed mapping of dropout and mobility dynamics among Italian university students in the 2016–2025 period. The analysis will make it possible to distinguish between different student trajectories that are often treated as a single phenomenon, such as permanent exit from the university system, transfer to another university, change of degree programme, graduation, or movement towards online education.

The results will contribute to the empirical evidence needed to monitor the Italian higher education system and to design more effective dropout–prevention policies. In particular, the study will provide

insights into the stages of university careers in which students are more likely to leave, move or complete their studies, and into the individual and contextual factors associated with these trajectories. By linking students' university careers with their educational background at entry, the analysis will also contribute to a better understanding of school–to–university transitions and of the conditions that may support successful participation and completion in higher education.

Keywords: Student dropout; Academic trajectories; Survival analysis; Administrative education data

Insights from Italy's Traditional and Online University Students

Andrea Priulla – Cristina Usala – Gabriele Lombardi

Introduction

Over the last two decades, online education has moved from being a marginal innovation to becoming a structural component of contemporary higher education systems. Early contributions on digital learning interpreted the diffusion of computer–mediated communication as the beginning of a profound transformation of higher education, potentially replacing the traditional, classroom–based model with a more flexible, open, and networked educational paradigm (Harasim, 2006). Although online education has not replaced traditional face–to–face programmes, its relevance has increased steadily across countries and institutional settings (Cheslock & Jaquette, 2022).

Italy represents a particularly relevant case for studying this transformation. Online universities were formally introduced in the early 2000s and have expanded substantially over the last decade. Their educational offer has more than doubled, while the number of enrolled students has increased by more than 400%. As a result, students enrolled in online universities now account for more than 10% of the Italian university population (Marra, 2024). This rapid expansion has changed the structure of the national higher education system and generated intense debate about the quality, regulation, and social function of online institutions.

Online universities occupy an ambiguous position in the Italian higher education system. On the one hand, their flexibility and lower participation costs may broaden access for non–traditional students, including older, working, or geographically constrained individuals, thereby helping to address Italy's low tertiary attainment and persistent dropout problem (EUROSTAT, 2025). On the other hand, their expansion has raised concerns about academic quality, limited student–teacher interaction, weaker social integration, and market–oriented organisational models. Since online students differ substantially from traditional students, evaluating their role requires empirical evidence rather than a purely normative assessment.

Recent empirical evidence has contributed to this debate by estimating the average effect of online university attendance on students' academic outcomes in Italy. Using administrative data and propensity score matching, Usala et al. (2026) show that students enrolled in online universities are, on average, less likely to drop out after the first year than comparable students enrolled in traditional universities. The retention effect is particularly pronounced among students who enrol later in life, those from technical and vocational schools, and those from Southern Italy. These findings suggest that online universities may not have a uniform effect across the student population.

Building on this evidence, the present paper addresses this question by shifting the focus from average treatment effects to treatment–effect heterogeneity. To this end, this paper adopts a causal forest approach to estimate Conditional Average Treatment Effects (CATEs) of attending an online university rather than a traditional one. In practical terms, the approach provides an estimated treatment effect for each student, conditional on observed characteristics, including sociodemographic information and high school and university career variables. This approach allows

us not only to ask whether online attendance reduces dropout on average, but also to identify student profiles for whom online universities may be an effective alternative. Understanding this distinction is crucial for informing public policy on accreditation, regulation, quality assurance, and student support in Italian higher education.

Data

We use data from the MOBYSU.IT database containing administrative information on the academic careers of the university student population between 2010 and 2019. We exclude cohorts enrolled after 2019 because the COVID–19 pandemic exposed both traditional and online students to some form of remote teaching, potentially blurring the distinction between the two types of institutions.

MOBYSU.IT also provides a set of individual, educational, and geographical characteristics, including: sex, age at enrolment, year of enrolment, high–school final mark, regularity in high–school completion, high–school track, high–school macro–region, field of study, and the minimum travel distance between the student's city of residence and the nearest university. A key piece of information is the timing of university enrolment following high school graduation. Students who enrol immediately are classified as direct entrants, while those who enrol after a gap period are classified as non–direct entrants. This distinction is particularly relevant, as non–direct entrants account for a large share of the online student population (Usala et al., 2026).

To improve comparability between the two groups, we excluded those students enrolled in degree programmes not offered by any online institution and those with missing information on their high school background or city of residence, resulting in a final sample of 61,826 online students and 1,741,184 traditional students. Table 1 presents the distribution of the student population in our final dataset by academic outcome, distinguishing between students who enrolled immediately after high school graduation and those who enrolled later.

Tab. 1: Descriptive statistics of student population according to their academic outcomes.

Variables	Full Sample	Traditional Total	Online Total	Direct	Non–direct
No. observations	1,803,010	1,741,184	61,826	13,134	48,692
Academic outcomes					
Dropout	12.57%	12.37%	18.14%	21.26%	17.56%
Non dropout	87.43%	87.63%	81.86%	78.74%	82.44%

Methods

The aim is to investigate whether the effect of online university attendance on first–year dropout varies across student profiles. We define online students as the treated group, while traditional students constitute the control group. The outcome of interest is first–year dropout, defined as not being enrolled in any Italian university during the second year. All students who remain enrolled in the university system are classified as non–dropouts, including both those who continue in the same type of institution and those who switch between online and traditional universities.

We adopt a potential–outcomes framework and focus on Conditional Average Treatment Effects (CATEs), defined as:

$$\tau(x) = \mathrm{E}[Y_i(1) - Y_i(0) \mid X_i = x]$$

where $Y_i(1)$ and $Y_i(0)$are the potential dropout outcomes under online and traditional university attendance, respectively, and X_i is the vector of observed student characteristics. Since the outcome is binary, the CATE can be interpreted as the difference in the probability of dropout associated with attending an online rather than a traditional university for students with a given profile.

To estimate CATEs, we use causal forests (Wager & Athey, 2018). Building on the random forest algorithm introduced by Breiman (2001), this method extends traditional prediction–focused

approaches to the causal inference framework by identifying how treatment effects vary across individuals with different observed characteristics. Causal trees estimate CATEs by dividing the covariate space into subgroups that are as homogeneous as possible in treatment effects. Unlike standard regression models, this approach does not require us to pre–define all relevant subgroups or interaction terms. Instead, it allows heterogeneity in treatment effect to emerge flexibly from the data.

Before estimating the causal forest, we estimate two auxiliary components. First, we estimate the propensity score, that is, the probability that a student enrols in an online university given their observed characteristics. Second, we estimate the expected dropout probability conditional on the same covariates. This reflects that some students have a higher or lower baseline risk of dropping out, regardless of the type of university they attended. The causal forest then uses these two components to adjust the comparison between online and traditional students and to reduce the risk that differences in dropout are driven by pre–existing differences between the two groups.

The analysis is conducted separately by enrolment cohort. This allows both the selection into online universities and the effect of online attendance to vary over time, which is particularly important given the rapid expansion of online higher education in Italy between 2010 and 2019. In all models, we cluster observations by field of study, accounting for the fact that students enrolled in similar fields of study may share common academic environments, institutional features, and dropout dynamics. Since online students represent a relatively small share of each cohort, we use conservative forest specifications, with sufficiently large terminal nodes and splitting rules that limit excessive imbalance between treated and control students.

Finally, we study the distribution of the estimated CATEs across the student characteristics that emerge as most relevant from the causal forest variable–importance analysis. As shown in Figure 1, high–school final mark is the most important predictor of treatment–effect heterogeneity, followed by direct enrolment, type of high–school diploma and high–school macro–region.

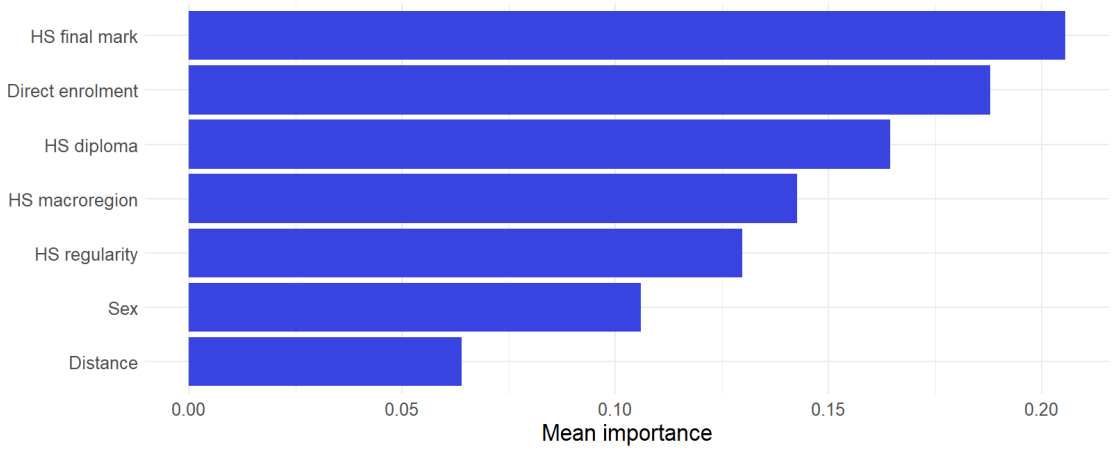


Fig. 1: Average variable importance across the cohort–specific causal forests. Higher values indicate variables that contribute more to explaining heterogeneity in the estimated treatment effects.

Preliminary results

Figure 2 shows that CATEs are negative for all groups, indicating that online enrolment is associated with a lower probability of dropout. The effect is stronger among students with lower high–school final marks, males, and non–direct entrants. The South and Islands stand out for both the magnitude and heterogeneity of the effects: among low–achieving students, the estimated dropout reduction ranges from about 10 percentage points for direct female entrants to around 16 percentage points

for non–direct male entrants. While effects in the North and Centre are generally smaller, roughly between 4 and 10 percentage points, in the South and Islands they remain sizeable even at higher levels of prior achievement, especially for male non–direct students.

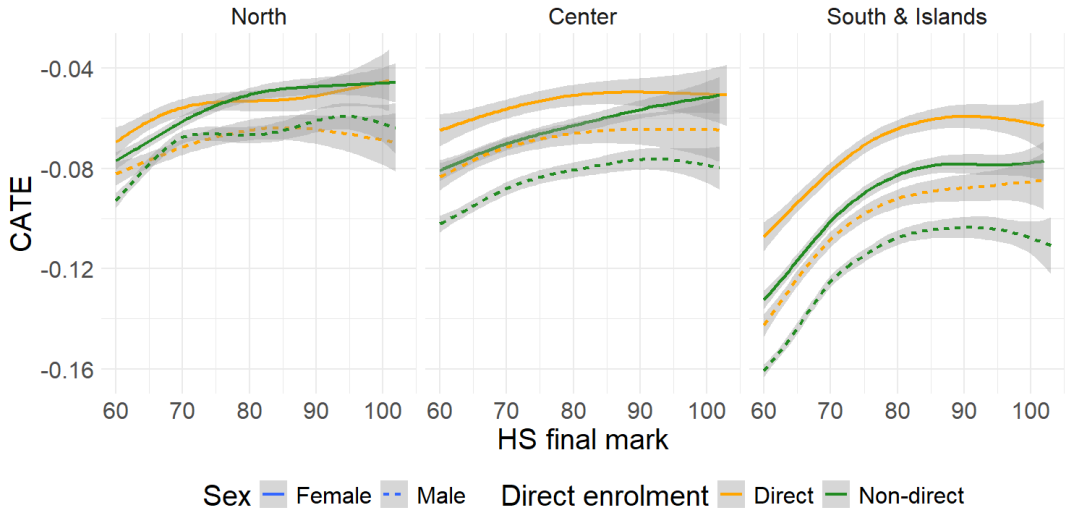


Fig. 2 – CATEs of online university attendance on dropout by high–school final mark, macro–region, sex, and direct enrolment at university

Conclusions

This paper investigates heterogeneity in the effect of online university attendance on first–year dropout in Italy. Building on previous evidence on the average retention advantage of online universities (Usala et al., 2026), we show that this effect is not uniform across students. Rather, the estimated CATEs suggest that online attendance is particularly associated with reduced dropout probabilities among students who may face greater barriers in traditional university settings (e.g., male, non–regular educational path, low academic performance). The heterogeneity of effects is also strongly structured by geography. Southern
students display larger heterogeneity in treatment effects, suggesting that online universities may play a more relevant compensatory role in areas where access to traditional higher education is more constrained. By contrast, patterns in the North and Centre appear more homogeneous and less pronounced. Overall, these findings suggest that online universities should not be evaluated solely by their average effect, but their role depends on students' educational and personal characteristics. In the Italian context, online education may serve as a complementary channel to reduce dropout among specific groups of students, especially those less likely to thrive in traditional university settings.

At the same time, the results call for caution. The analysis relies on observed characteristics and cannot fully account for unobserved factors such as employment status, motivation, family responsibilities, or preferences for flexible learning. Future research should further investigate these mechanisms and assess whether the retention advantage associated with online universities translates into longer–term academic success and labour–market outcomes.

Keywords: Higher education; Online universities; Academic outcomes; Dropout rates

SESSION 15. EDUCATIONAL POVERTY AND EQUITY IN EDUCATIONAL OPPORTUNITIES: MEASUREMENT, CONTEXTS, AND POLICY RESPONSES

ORGANIZERS: UNIVERSITY OF PIEMONTE ORIENTALE – UNIVERSITY OF MILANO BICOCCA

COORDINATORS: GIORGIA CASALONE – ALESSANDRA MICHELANGELI

5TH NOVEMBER: 4.30 P.M. – 6.30 P.M. {AULA A1.1 – RESEARCH 14}

Beyond Citizenship: Migration Background, School Dropout and Educational Inequalities in Italy

Lorenzo Maraviglia

Introduction

Educational inequalities associated with migration background pose a major challenge for contemporary education systems. In Italy, public and scientific debate has traditionally referred to citizenship (i.e. confronting "Italian" v. "Foreign" students) as the main criterion for describing educational outcomes among students with and without a migration background. Even popular indicators such as the Early Leavers from Education and Training (ELET) rate are now released according to citizenship and have contributed to shaping the perception of substantial educational disadvantages affecting foreign students.

However, statistics disaggregated by citizenship may be ambiguous. First, citizenship status is not stable over time. Each year, a considerable number of young people acquire Italian citizenship following the "italianisation" of their parents. As a consequence, individuals initially classified as foreign citizens may subsequently be included among Italians. Second, citizenship–based indicators combine highly heterogeneous populations, including students born and educated in Italy and individuals who arrived in the country only recently and therefore had limited exposure to the Italian education system.

These issues raise important questions regarding the measurement of educational inequalities. If citizenship changes over time and if naturalisation processes disproportionately involve families that are more socially integrated and educationally successful, then comparisons based solely on citizenship may supply a biased picture of inequalities affecting students with a migration background.

This paper contributes to the literature on educational inequalities by examining the extent to which the measurement of achievement gaps depends on the indicator adopted and on selective processes occurring throughout educational trajectories. Using longitudinal INVALSI data, the study investigates whether the reduction in achievement gaps between native students and students with a migration background observed between lower secondary and upper secondary education reflects genuine convergence in learning outcomes or is partly explained by selective school dropout.

Data Used

The study primarily relies on individual–level data collected through the Italian National Institute for the Evaluation of the Education System (INVALSI) assessments. We use four student cohorts that participated in Grade 8 assessments between 2016 and 2019 and can be followed until Grade 13 assessments conducted between 2021 and 2024.

INVALSI data provide detailed information on students' performance in Italian language and mathematics, as well as information on students' migration background. Unlike citizenship, migration background is based on place of birth and parental origin and therefore represents a substantially more stable characteristic for longitudinal analyses.

To contextualise the findings and assess the relationship between citizenship and migration background, the analysis also uses microdata from the Italian Labour Force Survey (ISTAT–RFL). These data allow us to compare patterns of educational attainment and early school leaving among native and migrant–background populations and to discuss the implications of citizenship acquisition for the measurement of educational inequalities.

Methodology

The paper adopts a descriptive longitudinal approach.

First, achievement gaps between native students and students with a migration background are calculated using INVALSI standardised scores in Italian and mathematics. Comparisons are conducted for every cohort both at Grade 8 and Grade 13, allowing the reconstruction of educational trajectories across different school levels.

Second, two distinct analytical strategies are implemented.

The first strategy compares the full populations observed at Grade 8 and Grade 13. This approach corresponds to the most common interpretation of achievement gaps and allows an assessment of how inequalities evolve over time when considering all students participating in each assessment cycle.

The second strategy focuses exclusively on students who participated in both Grade 8 and Grade 13 assessments. Restricting the analysis to this longitudinally observed sub–population makes it possible to isolate the effects of selective attrition and school dropout. Comparing results obtained through the two approaches allows us to estimate the extent to which changes in observable overall achievement gaps are associated with compositional changes in the student population.

Finally, evidence from the Labour Force Survey is used to examine differences in school dropout patterns between native students and students with a migration background and to interpret the mechanisms underlying the observed achievement trajectories.

Results and Discussion

The longitudinal analysis reveals a substantial stability of achievement patterns across cohorts.

For all four cohorts considered, students with a migration background perform significantly worse than native students in both Italian and mathematics at Grade 8. Achievement gaps are consistently larger in Italian than in mathematics.

When comparing the full populations observed at Grade 8 and Grade 13, the achievement gap appears to decrease substantially over time. For example, in the cohort assessed in Grade 8 in 2016 and Grade 13 in 2021, the gap in Italian decreases from –23.2 to –13.4 points on the common standardised scale, corresponding to a reduction of nearly ten points. Similar patterns are observed across all cohorts and in both subject areas.

	Observed gap between native students and with migration background: Grado_8		Observed gap between native students and with migration background: Grado_13		Reduction of gap from 8th to 13th grade	
Longitudinal cohort	Ita	Mat	Ita	Mat	Ita	Mat
Cohort 2016 (8°) 2021 (13°)	–23,2	–17,6	–13,4	–7,9	9,8	9,7
Cohort 2017 (8°) 2022 (13°)	–24,3	–16,1	–13,8	–7,5	10,5	8,6
Cohort 2018 (8°) 2023 (13°)	–24,9	–13,0	–12,4	–6,0	12,5	7,0
Cohort 2019 (8°) 2024 (13°)	–22,6	–11,8	–13,6	–7,0	9,0	4,8

Source: INVALSI data (our elaboration).

At first sight, these findings could be interpreted as evidence of the capacity of the Italian education system to reduce inequalities associated with migration background. However, a different picture emerges when the analysis is restricted to students observed at both time points.

For the 2016–2021 cohort, the achievement gap in Italian among students who remain observable throughout the educational trajectory decreases from –16.2 (8th grade) to –13.3 points (13th grade), corresponding to a reduction of approximately three points only. A similar pattern is observed in mathematics, where the reduction in the gap falls from nearly ten points in the full–population comparison to approximately five points in the longitudinally matched sample. Comparable dynamics emerge for the remaining cohorts.

These findings suggest that a substantial share of the apparent reduction in educational inequalities is attributable to selective processes rather than to genuine convergence in learning outcomes. Students with a migration background are more likely to leave the education system before reaching the final year of upper secondary school, resulting in a progressively more selected population among those who remain enrolled.

Evidence from ISTAT Labour Force Survey data supports this interpretation. Students with a migration background exhibit rates of early school leaving that are approximately from two to two and a half times higher than those observed among native students. Consequently, analyses based exclusively on average achievement among students who remain in education risk underestimating the extent of educational disadvantage and overestimating the degree of convergence between groups.

The results also highlight the limitations of citizenship–based indicators. Since citizenship acquisition removes a selective segment of students from the foreign population and reclassifies them as Italian citizens, analyses based on citizenship may further contribute to a distorted respresentation of educational inequalities associated with migration background.

Conclusions

This paper highlights the importance of measurement choices in the study of educational inequalities.

The findings suggest that migration background constitutes a relevant indicator for analysing educational trajectories and inequalities over time. Unlike citizenship, migration background is a stable characteristic and is therefore less vulnerable to distortions generated by naturalisation processes.

The analysis further demonstrates that a substantial proportion of the observed reduction in achievement gaps between native students and students with a migration background can be explained by selective school dropout. As a result, improvements in average achievement should be interpreted cautiously, particularly when they occur alongside substantial differences in educational participation and completion.

More broadly, the study emphasises the need to analyse educational outcomes and educational persistence jointly. Failure to consider both dimensions may lead to misleading conclusions regarding the capacity of education systems to reduce inequalities and promote equality of opportunity.

Future research would benefit from a stronger integration of educational and demographic data sources, allowing researchers to jointly analyse citizenship, migration background, educational trajectories, and school completion patterns. Such integration would provide a more comprehensive understanding of educational inequalities and support the development of more effective evidence–based policies.

Keywords: Educational inequalities; Migrant background; School dropout; INVALSI

Educational Poverty and 21st Century Skills Development: A Quasi–Experimental Evaluation of the UPSHIFT Programme in Italy

Sabrina Lipari

Introduction

In recent years, educational poverty has emerged as one of the most pressing challenges for the contemporary education systems, gaining increasing prominence in both academic discourse and public policy agendas. In Italy, this phenomenon is multidimensional, as it is closely intertwined with economic, cultural, and social conditions that profoundly shape learning opportunities and human capital development. Despite notable progress in recent years, the education system continues to exhibit significant territorial and socio–economic disparities, which both reflect and reproduce broader societal inequalities[13].

According to the most recent estimates, the early school leaving rate stands at approximately 8–9% in the 2025–2026 period. This indicator is aligned with, or slightly below, the European average suggesting significant long–term improvement, although persistent critical issues remain across regions and social groups[14]. These disparities are particularly evident in Southern Italy and among students with a migrant background, for whom dropout risks can exceed 25–26%, underscoring a strong stratification of educational opportunities[15].

Alongside explicit school dropout, increasing attention has been paid to the phenomenon of implicit dropout, which concerns students who complete their education but fail to achieve adequate proficiency in basic competencies. According to the 2025 INVALSI Report, approximately 12.3% of lower secondary school students and 8.7% of upper secondary school students perform below the minimum expected proficiency levels. These data reveal the existence of a substantial proportion of young people exposed to a less visible yet highly detrimental form of educational vulnerability, carrying profound implications for equity.

Regarding learning outcomes, a systemic challenge also emerges: the results from OECD–PISA 2024 indicate below–average performance in several domains, such as creative thinking (31 points compared to the OECD average of 33), thereby signaling weaknesses in the development essential competencies required to navigate complex environments. These dynamics are further reflected in school–to–work transitions: youth unemployment remains around 21–22%, while the share of NEETs (Not in Education, Employment or Training) ranges between 15% and 16%, with higher rates observed in Southern Italy[16]. Literature consistently highlights a strong correlation between low proficiency levels, school dropout, and the risk of socio–economic exclusion[17].

Within this framework, cultural capital — defined as the availability of educational resources and participation in cultural experiences — plays a pivotal role in academic achievement and the acquisition of transversal skills, thereby either reinforcing or mitigating educational inequalities[18].

Against this backdrop, the *UPSHIFT* programme — promoted by *UNICEF* and *Junior Achievement Italy* in collaboration with the Italian Ministry of Education — aims to reduce educational poverty and

---

[13] ISTAT (2026). La povertà educativa e disuguaglianze territoriali.

[14] Eurostat (2025). Early leavers from education and training (ELET).

[15] INVALSI (2025). Rapporto nazionale 2025.

[16] SOS Italia, citato in Human Foundation, *UPSHIFT - Report di valutazione 2025*, p. 7, https://www.humanfoundation.it/2022/wp-content/uploads/2025/11/Upshift-2025_Report-valutazione_Novembre.pdf.

[17] Pierpaolo Triani, citato in Human Foundation, *UPSHIFT - Report di valutazione 2025*, p. 7, https://www.humanfoundation.it/2022/wp-content/uploads/2025/11/Upshift-2025_Report-valutazione_Novembre.pdf.

[18] Bourdieu, P. (1979). La distinction: Critique sociale du jugement. Paris: Les Éditions de Minuit.

foster 21st–century skill development among lower and upper secondary school students, with a specific focus on disadvantaged contexts.

Its theoretical framework integrates the *EntreComp*[19] and *LifeComp*[20] models alongside the *UNICEF transferable skills framework*[21], proposing a holistic approach to learning that combines cognitive, socio–emotional, and entrepreneurial dimensions. Aligning with these approaches, UPSHIFT adopts active and experiential teaching methodologies — such as *learning–by–doing* and *design thinking* — which literature recognizes as highly effective for transversal skill development[22]. However, evidence shows that the effectiveness of such interventions is strongly mediated by students' socio–economic and cultural backgrounds, underscoring the need for careful analysis of their differential effects.

In light of these considerations, this study analyzes the impact of the UPSHIFT programme on the development of 21st–century skills, focusing particularly on how cultural capital and contextual conditions modulate educational outcomes. In doing so, this paper contributes to the academic debate on educational poverty and policies aimed at fostering equity within the education system.

Data used

The analysis is based on data collected during the evaluation of the *UPSHIFT* programme for the 2023–2025 period, utilizing a nationwide sample of lower and upper secondary school students.

Schools were selected based on socio–educational eligibility criteria defined in collaboration with institutional partners (UNICEF and the Italian Ministry of Education). Specifically, priority was given to public schools located in vulnerable contexts characterized by low academic achievement (e.g., INVALSI test results), a high risk of early school leaving, a significant presence of students with a migratory background, and socioeconomic disadvantage. This approach allows the analysis to focus on environments that are highly relevant to the study of educational poverty.

The sample comprises several hundred students across different educational levels (lower secondary schools – *UPSHIFT for Teens* and upper secondary schools – *UPSHIFT for Youth*). This sample is geographically diverse — spanning the North, Center, South, and Islands—and heterogeneous in terms of students' socioeconomic backgrounds. Such demographic variety reflects a wide range of educational settings, allowing for stratified analyses of key variables, including cultural capital, cultural participation, and the risk of educational poverty.

A pre–post longitudinal design was adopted for data collection, administering standardized digital questionnaires at two specific time points: pre–intervention (T1) and post–intervention (T2). The evaluation instruments, grounded in the European *EntreComp* and *LifeComp* frameworks, incorporate self–assessment scales targeting five core dimensions:

- life skills;
- entrepreneurial competencies;
- self–efficacy and future orientation;
- cultural participation;
- objectified cultural capital (i.e., access to material resources for learning).

The scales demonstrate strong internal reliability, as evidenced by Cronbach's alpha coefficients exceeding the acceptable threshold ($\alpha > 0{,}70$), thereby validating the psychometric robustness of the instruments. Broadly, this dataset structure allows for evaluating both post–intervention changes and outcome variances across different background contexts, facilitating an integrated analysis of intervention efficacy in relation to students' baseline conditions.

[19] European Commission (2016). EntreComp: The European Entrepreneurship Competence Framework.

[20] European Commission (2020). LifeComp: The European Framework for Personal, Social and Learning to Learn Key Competence.

[21] UNICEF (2019). Global Framework on Transferable Skills.

[22] OECD (2010). The Nature of Learning: Using Research to Inspire Practice; Kolb, D. (1984). Experiential learning: experience as the source of learning and development.

Methodology

A quasi–experimental, pre–test/post–test design was adopted for the treated group[23]. Despite the absence of a control group, this configuration enables the detection of competency variations attributable to the intervention. Framed within a *theory–based evaluation* paradigm, the methodology references an explicit *Theory of Change* to link four core dimensions:

- *inputs* (resources and activities);
- *outputs* (programme participation);
- *outcomes* (competency development);
- *anticipated impacts* (empowerment and social inclusion).

Quantitative analysis proceeds through pre–post mean comparisons, sub–construct evaluations, and result stratification across contextual variables, including cultural capital, cultural participation, and educational poverty risk. Particular emphasis is placed on uncovering differential impacts to understand the programme's role in mitigating or exacerbating educational disparities.

4.Results and discussion

Findings indicate a statistically significant improvement in students' competencies following the *UPSHIFT* intervention, confirming its efficacy in cultivating transversal skills within heterogeneous educational settings. The aggregate life skills index rose from 4.3 to 4.5 in upper secondary schools and from 4.4 to 4.5 in lower secondary institutions.

Component–specific analysis reveals that the most substantial shifts occurred in:

- emotion regulation (+0.5);
- stress management (+0.4);
- problem–solving/critical thinking (+0.2/+0.3).

This evidence resonates with experiential pedagogy literature highlighting the capacity of learning–by–doing frameworks to catalyze socio–emotional growth[24].

Parallel advancements were documented in entrepreneurial competencies, where mean scores advanced from 4.1 to 4.3 (upper secondary) and from 4.3 to 4.6 (lower secondary). Prominent gains clustered around:

- planning and management (+0.3, reaching +0.6 among the youngest cohort);
- motivation and perseverance (+0.3);
- alongside risk mitigation and goal–directed collaboration (+0.2).

Conversely, a marginal contraction in self–efficacy (–0.1 in select groups) likely reflects a shift toward more grounded self–evaluation.

Stratification underscores the critical influence of baseline conditions, while students with substantial cultural capital exhibited increases up to +0.4 in life skills and +0.2/+0.3 in entrepreneurship, those in resource–constrained cultural contexts displayed attenuated growth or recalibration trends (e.g., upper secondary entrepreneurship scores shifting from 3.9 to 3.3).

Active cultural participation likewise predicted superior post–test scores (up to 4.7–4.8), corroborating the foundational role of cultural assets. Notably, the most pronounced effects emerged at the lower secondary level, suggesting that early interventions maximize skill acquisition.

In summary, this evaluation[25] yields threefold conclusions:

---

[23] The evaluation phase was conducted by Human Foundation, an independent research organization specializing in social impact assessment, within the framework of a study commissioned by Junior Achivement Italia.

[24] Faella, E., et al. (2025). Educational practices in motion: A scoping review of embodied learning approaches in school. Frontiers in Education, 10, articolo 1568744.doi.org

[25] The reported mean variations were derived from the analysis of pre- and post-intervention survey data features in the *2025 UPSHIFT Evaluation Report* developed by the Human Foundation. These variations correspond to the sections focusing on life skills (pp. 29–31) and entrepreneurial competencies (pp. 34–35),

- the UPSHIFT model substantially enhances 21st–century competencies;
- programme outcomes are heavily conditioned by structural variables (cultural capital and participation);
- educational interventions exhibit greater elasticity during earlier developmental stages.

Ultimately, these data support integrating experiential learning with targeted, personalized strategies to promote genuine educational equity

Conclusions

This research confirms that the UPSHIFT programme serves as an effective, innovative catalyst for expanding 21st–century competencies and addressing educational deprivation. However, the data indicate that overarching programme effectiveness is strongly conditioned by students' socio–cultural context. Specifically, baseline cultural capital and cultural participation act as critical determinants of skill development, validating the demand for context–sensitive pedagogical designs. Based on these empirical insights, the following policy interventions are recommended:

- prioritizing early–stage interventions within lower secondary educational tracks;
- structuring individualized support frameworks tailored to vulnerable student cohorts;
- incentivizing access to extracurricular learning environments and cultural capital assets;
- consolidating experiential and collaborative learning modalities within core institutional curricula.

Beyond these immediate implications, the evaluation has informed the structural refinement of the UPSHIFT model itself, demonstrating its inherent capacity for adaptation and scaling. The evolution of the UPSHIFT programme further reflects a progressive process of institutionalization within the Italian education system. Initially developed as an innovation intervention (2014) and introduced in Italy in 2018, UPSHIFT has progressively been integrated into broader policy frameworks, including the European Child Guarantee (2020), institutional agreements with the Ministry of Education and Merit (2022), and collaborations with INDIRE and the "*Avanguardie Educative*" movement (2023). More recently, the programme has entered a consolidation phase within formal education pathways (2024–2025), including its integration into civic education, teacher training, and school–to–work transition initiatives. This trajectory highlights the potential of the programme to transition from a project–based intervention to a structurally embedded, policy–relevant educational model.

A notable development is the 2025–2027 implementation of UPSHIFT Generazione Pari, developed by Junior Achievement Italy in collaboration with UNICEF and Fondazione CDP, which explicitly focuses on gender equity, the prevention of gender–based violence, self–entrepreneurship, skill acquisition, and civic engagement.

In conclusion, sustainable educational reform requires moving beyond episodic interventions toward the institutionalization of effective pedagogical innovations. This entails integrating experiential and competence–based approaches within formal curricula and education policy frameworks. Such a shift requires a systemic and multi–level perspective capable of combining methodological innovation, educational equity, and long–term sustainability.

Only through this form of structural embedding can high–impact educational opportunities be equitably expanded across socio–educational contexts. Future research should further explore the long–term impact and scalability of such models across different educational systems.

Keywords: Educational poverty; 21st century skills; Cultural capital; Quasi–experimental evaluation

as well as the detailed analyses of cultural capital and cultural participation (pp. 32–33, 36–37). humanfoundation.it/2022/wp-content/uploads/2025/11/Upshift-2025_Report-valutazione_Novembre.pdf

Breaking the cycle? Educational poverty persistence among italian high school students

Maria Berrittella – Francesco Salomone Marino

Educational poverty is generally defined as the inability of the individuals to acquire the fundamental cognitive, social, and educational skills necessary for full participation in society. Educational poverty can be also interpreted as a condition of unequal educational opportunities, reflecting disparities in access to resources, learning environments, and developmental experiences. From this perspective, educational poverty extends beyond differences in schooling achievement levels and encompasses a multidimensional phenomenon shaped by factors that are largely outside students' direct control, including family socioeconomic background, school characteristics and territorial contexts.

Focusing on Italian high school students, this paper pursues two main aims. First, it estimates the extent to which educational poverty persists across generations, examining the relationship between parents' education and their descendant's educational outcome, measured , for example, using INVALSI test scores. Second, it investigates the role played by school characteristics and territorial disparities in shaping the educational trajectories of students who experience high levels of educational poverty persistence. To address these aims, we employ multinomial logistic regression models using longitudinal data from the MOBYSU.IT database (2020), an integrated dataset that combines information from the National Archive of Schools, the database of the National Evaluation Institute for the School System, and the National Archive of University Students.

The analysis reveals a strong and persistent association between students' socioeconomic background and their academic achievement. High school students from families characterized by lower levels of parental education and weaker economic conditions are significantly more likely to achieve lower educational outcomes and to remain trapped in low–performance trajectories throughout their schooling. These findings point to substantial intergenerational persistence in educational disadvantages and limited opportunities for upward educational mobility.

At the same time, school performance and school–level characteristics emerge as crucial factors in either reinforcing or mitigating educational poverty. Schools serving socioeconomically disadvantaged populations tend to exhibit lower average academic achievement, a higher concentration of low–performing students, and fewer institutional resources capable of compensating for pre–existing inequalities. Nevertheless, the results show that attendance at high–performing upper secondary schools can partially reduce the disadvantages associated with an unfavorable family background, highlighting the potential role of schools as mechanism of educational inclusion and social mobility.

Territorial inequalities represent an additional and highly relevant dimension of educational poverty persistence. Significant differences in educational outcomes are observed across Italian regions, with marked achievement gaps between Northern and Southern Italy. These disparities reflect broader socioeconomic imbalances, unequal access to educational resources and services, differences in local labor market opportunities, and varying levels of institutional effectiveness. The territorial context therefore acts not merely as a background condition, but as an active mechanism influencing students' educational opportunities and achievements.

Overall, the findings suggest that educational poverty is a multidimensional, cumulative, and persistent phenomenon generated by the interaction of individual, family, school, and territorial factors. Family disadvantages, weak school performance, and deprived territorial contexts mutually reinforce one another, producing mechanisms that reproduce educational inequalities across generations. The study contributes to the literature on educational inequality by providing empirical evidence on the intergenerational transmission of educational poverty in Italy and by highlighting

the importance of integrated policy interventions aimed at simultaneously addressing family vulnerabilities, improving school quality, and reducing territorial disparities. Such policies are essential to promote equal educational opportunities and to break the cycle of educational poverty.

Keywords: Education poverty persistence; Family background; High school performance; INVALSI test scores; Upper Secondary education.

Who Falls Behind? Educational Inequality Across Places and Social Groups in Italy

Alessandra Michelangeli – Giorgia Casalone – Kateryna Tkach

Introduction

Education is widely recognized as a fundamental determinant of individual well–being, social mobility, and long–term economic development. Despite its pivotal role in fostering equal opportunities, educational opportunities and outcomes remain unevenly distributed across socio–economic groups and territories. Understanding how these disparities evolve over time is therefore crucial for designing effective educational and social policies.

The literature on educational inequality identifies several factors contributing to unequal learning outcomes, including family background, parental education, migration status, school quality, and territorial characteristics (Brunori et al. 2026; Pensiero et al. 2019). In Italy, despite a highly centralized education system, significant regional and local disparities in student achievement persist. Previous studies have documented marked territorial differences in educational attainment and educational poverty, often associated with broader socio–economic inequalities (Pratesi et al., 2021; Cafieri et al. 2026).

This study investigates changes in educational inequality in Italy over the period 2017–2023, encompassing the COVID–19 pandemic. The educational disruptions associated with school closures and distance learning generated substantial learning losses, particularly among disadvantaged students, and contributed to widening educational disparities documented both internationally (OECD, 2021; World Bank et al., 2021) and in Italy (Borgonovi & Ferrara, 2023; Bertoletti et al., 2023, Casalone et al., 2023).

Building on the literature on inequality measurement and spatial disparities, we adopt the Generalized Entropy (GE) class of inequality indices, focusing on the Theil index because of its additive decomposability (Theil, 1967; Shorrocks, 1980). This framework allows us to distinguish between inequality arising within socio–demographic groups and territorial units and inequality attributable to differences between them.

The paper contributes to the literature in three ways. First, it provides a longitudinal analysis of educational inequality using individual–level INVALSI data covering the period 2017–2023. Second, it examines territorial disparities at the provincial (NUTS–3) level, allowing a detailed assessment of spatial inequalities. Third, it explores how educational inequality differs according to students' gender, migration background, and parental educational attainment, thereby highlighting the intersection between territorial and socio–economic dimensions of disadvantage.

The remainder of this extended abstract is structured as follows. The next section introduces the data, followed by the methodological framework, the empirical results, and the concluding remarks.

Data

The empirical analysis relies on individual–level data from the National Institute for the Evaluation of the Education and Training System (INVALSI). The database includes standardized test scores in

Italian language and Mathematics for students enrolled in Grade 2 (primary school), Grade 5 (primary school), and Grade 8 (lower secondary school).

The analysis covers five academic years: 2017–2018, 2018–2019, 2020–2021, 2021–2022, and 2022–2023. The period therefore spans both pre–pandemic and post–pandemic phases, making it possible to assess how school closures and distance learning were associated with changes in educational inequality.

Student achievement is measured through WLE (Weighted Likelihood Estimation) scores standardized by INVALSI according to the Rasch scaling methodology. The dataset also contains information on students' gender, immigration status, parental educational attainment, and province of residence. Immigration status distinguishes native students, first–generation immigrants, and second–generation immigrants. Parental education is classified according to whether at least one parent possesses a university degree.

The provincial dimension is particularly important because it captures territorial heterogeneity in socio–economic conditions, educational resources, and local opportunities. Average test scores reveal a persistent geographical pattern, with higher educational performance generally concentrated in northern provinces and lower outcomes observed in several southern areas. These territorial disparities appear particularly pronounced at the lower secondary school level.

Methodology

Educational inequality is measured using the Generalized Entropy family of indices. Specifically, the analysis employs the Theil index, which is especially suitable because it can be decomposed into within–group and between–group components.

Let individual student achievement be represented by standardized test scores. The overall Theil index measures the dispersion of outcomes around the national average. Following Shorrocks (1980), total inequality can be decomposed into:

- a within–group component, representing disparities among students belonging to the same group;
- a between–group component, representing differences in average achievement across groups.

The decomposition is applied using several grouping variables:

1. Province (NUTS–3 level);
2. Gender;
3. Immigration status;
4. Parental educational attainment.

Additionally, inequality is decomposed by province within selected subpopulations, namely female students, first–generation immigrant students, and students with at least one university–educated parent.

The analysis is complemented with panel regression models estimated at the provincial (NUTS–3) level. The dependent variable is the Theil index of educational achievement, calculated separately for Italian language and Mathematics test scores. The regressions include province, year, and grade fixed effects, allowing us to identify changes in educational inequality over time while controlling for persistent territorial characteristics.

Results

The descriptive analysis reveals a marked increase in educational inequality following the COVID–19 pandemic, particularly among primary school students. The decomposition of the Theil index shows that inequality is driven predominantly by within–group disparities rather than by differences between provinces or socio–demographic groups. Nevertheless, significant territorial clustering of educational outcomes remains evident, confirming the persistence of spatial inequalities across Italy.

The regression results provide additional insights into the determinants of educational inequality. First, educational inequality is significantly lower in Grades 5 and 8 compared with Grade 2, suggesting that disparities are most pronounced at the beginning of the educational cycle. For both

Italian and Mathematics, the coefficients associated with Grade 5 and Grade 8 are negative and highly significant.
Second, the temporal coefficients reveal a clear post–pandemic increase in inequality. Relative to the pre–pandemic academic year 2017–2018, inequality rose significantly in 2021–2022 and increased even further in 2022–2023. The effect is observed for both subjects and remains robust across model specifications, providing evidence that the consequences of the pandemic persisted well beyond the reopening of schools.
Third, socio–demographic composition appears to matter. Provinces with a larger share of second–generation immigrant students tend to display lower levels of educational inequality in both Italian and Mathematics. The coefficients are negative and statistically significant, suggesting that second–generation immigrants may be more integrated into local educational systems than first–generation students. By contrast, the share of first–generation immigrant students is not significantly associated with educational inequality.
The results also indicate that the proportion of highly educated parents is not significantly related to provincial educational inequality once territorial and temporal fixed effects are included. Similarly, gender composition shows limited explanatory power, with only weak evidence of lower inequality in Mathematics in provinces with a higher proportion of female students.
Finally, the interaction between educational level and the post–COVID period suggests that the pandemic may have affected educational cycles differently. While the interaction term is generally not significant for Grade 5, there is evidence that inequality in Mathematics increased more strongly among Grade 8 students during the post–pandemic period.
Taken together, these findings indicate that COVID–19 amplified existing educational disparities, with particularly persistent effects among younger students. At the same time, the results confirm that educational inequality cannot be explained solely by territorial differences; substantial heterogeneity remains within provinces and across socio–demographic groups.

Conclusions

This paper investigates the evolution of educational inequality in Italy before and after the COVID–19 pandemic using INVALSI standardized test scores and decomposable inequality measures.
Three main findings emerge. First, educational inequality increased significantly following the pandemic, especially in the years immediately after the return to in–person schooling. Second, inequality is driven primarily by disparities within provinces rather than by differences between provinces, highlighting the importance of individual and family–level factors. Third, the regression analysis reveals that inequality is particularly pronounced in the early stages of schooling and that post–pandemic increases persisted even after controlling for territorial fixed effects and demographic composition.
From a policy perspective, the results suggest that interventions aimed at reducing learning inequalities should focus not only on disadvantaged territories but also on vulnerable groups within territories. The persistence of post–pandemic effects points to the need for targeted remedial programs, particularly in primary education where learning losses appear to have translated into greater educational inequality. More broadly, strengthening the resilience of educational systems requires policies capable of mitigating the unequal consequences of future shocks and ensuring equitable learning opportunities across both places and social groups.

Keywords: Educational inequality; territorial disparities; learning outcomes

Tracking Equality of Opportunity over Compulsory Schooling: Evidence from Longitudinal INVALSI data

Giorgia Casalone – Alessandra Michelangeli

Introduction

Equality of Opportunity (EOp) in education requires that educational outcomes should not depend on circumstances beyond individuals' control, such as family background, parental education, socioeconomic status, gender, migration background, or territorial conditions. While a large empirical literature has investigated educational inequality of opportunity, considerably less attention has been devoted to understanding whether educational inequalities persist and accumulate throughout students' educational trajectories.

The empirical analysis builds upon the Equality of Opportunity framework developed by Roemer (1998) and subsequently operationalized by Bourguignon et al. (2007), Ferreira and Gignoux (2011), and Checchi and Peragine (2010). The EOp approach has been increasingly applied to educational achievement and educational attainment, documenting the persistent role of family and contextual factors in shaping educational opportunities (Schütz et al., 2008; Gamboa and Waltenberg, 2012; Ferreira and Gignoux, 2014; Brunori et al., 2026; Palmisano et al., 2022). However, most existing studies adopt a cross–sectional perspective and provide a static assessment of educational inequality.

This paper contributes to the literature by adopting a longitudinal perspective and investigating how inequality of opportunity evolves and accumulates throughout compulsory schooling. More specifically, the paper: (i) estimates the role of family and territorial circumstances at different educational stages; (ii) investigates whether family and territorial circumstances continue to influence educational outcomes after controlling for prior achievement; and (iii) develops a longitudinal application of Roemer's types–and–tranches approach to analyze whether students exhibiting similar levels of effort experience different educational trajectories depending on their family background (TBD).

Data

The empirical analysis uses longitudinal data from the Italian National Institute for the Evaluation of the Education System (INVALSI). Specifically, we follow the same cohort of students from Grade 2 in the 2018/2019 school year to Grade 5 in 2021/2022 and subsequently to Grade 8 in 2024/2025.

The longitudinal structure of the data allows us to analyze educational inequality across two critical transitions in students' educational careers. The first transition (Grade 2 to Grade 5) captures the progression through primary school, while the second transition (Grade 5 to Grade 8) encompasses the transition from primary to lower secondary education, a stage characterized by substantial changes in curricula, teaching practices, peer groups, and educational environments.

The dataset includes standardized test scores in mathematics and reading, as well as detailed information on students' family background, migration status, parental education, parental occupation, and geographical location.

Methodology

In general terms, educational achievement can be represented as $Y_i=f\,(C_i, E_i, \varepsilon_i)$ where Y_i denotes educational achievement, C_i is a vector of predetermined circumstances, E_i represents effort and other responsibility factors, and ε_i captures unobserved determinants.

Following the ex–ante approach to Equality of Opportunity, we first estimate the contribution of individual and family circumstances to educational achievement through the cross–sectional specification:

$$Y_{ig} = \alpha + \beta' C_i + \varepsilon_i \quad (1)$$

where g denotes the educational stage (Grade 2, Grade 5, and Grade 8), and C_i includes gender, migration background, parental education, and parental occupational status. The coefficient of determination, R^2_g, is interpreted as a lower–bound estimate of educational inequality of opportunity as some relevant circumstances remain unobserved.
To account for territorial inequalities, we estimate:

$$Y_{ig} = \alpha + \beta' C_i + \lambda_p + \varepsilon_i \quad (2)$$

where λ_p denotes province fixed effects. The comparison between equations (1) and (2) allows us to assess the relative contribution of family and territorial circumstances to educational inequality.
To investigate the dynamic evolution of educational inequality, we exploit the longitudinal structure of the INVALSI data. For each educational transition, we estimate:

$$Y_{ig+3} = \alpha + \beta\, Y_{ig} + u_i \quad (3)$$

And

$$Y_{ig+3} = \alpha + \beta\, Y_{ig} + \gamma'\, C_i + u_i \quad (4)$$

Dynamic inequality of opportunity is measured as:

$$EOp^{dynamic} = R^2(\, Y_{ig+3} \mid Y_{ig},\, C_i) - R^2(\, Y_{ig+3} \mid Y_{ig},) \quad (5)$$

where Y_{ig} denotes educational achievement at the previous educational stage, Y_{ig+3} the educational achievement three grades later.
To distinguish inequalities operating between and within schools, we estimate school fixed–effects models:

$$Y_{ig+3} = \alpha + \beta\, Y_{ig} + \gamma'\, C_i + \delta_s + u_i \quad (6)$$

where δ_s denotes school fixed effects. This specification identifies the effect of family background among students attending the same school.
We further investigate whether family background affects the process of skill accumulation through interaction models between prior achievement and parental education:

$$Y_{ig+3} = \alpha + \beta\, Y_{ig} + \theta(Y_{ig} \times ESCS_i) + \gamma'\, C_i + \delta_s + u_i \quad (7)$$

A positive and significant coefficient θ provides evidence of cumulative advantage in educational achievement.
As an extension of the standard ex–ante Equality of Opportunity framework, we are developing a longitudinal application of Roemer's types–and–tranches approach. Students are grouped into "types" defined by predetermined circumstances, while their relative position within the type–specific achievement distribution in Grade 2 is used as a proxy for effort. Subsequent educational outcomes are then compared across types for students occupying the same quantile (tranche), allowing us to investigate whether educational trajectories differ among students exhibiting similar levels of effort.

Preliminary results and discussion

Preliminary evidence reveals a statistically significant but quantitatively limited role of observed circumstances throughout compulsory education (Tables 1 and 2). At the national level, family circumstances explain between 6% and 7% of the variance in achievement in Grade 2, around 7% in

Grade 5, and between 12% and 15% in Grade 8. Family background appears to play a slightly stronger role in reading than in mathematics, particularly at later educational stages.
The analysis also reveals substantial geographical heterogeneity. Family circumstances explain a larger share of achievement variation in Northern Italy than in the Centre and South, whereas territorial circumstances play a stronger role in Southern regions, particularly during lower secondary education.

Table 1. Share of Variance in Mathematics Achievement Explained by Family and Territorial Circumstances, by Geographical Area and Educational Stage

	Grade 2			**Grade 5**			**Grade 8**	
Area	Family	Family + Territory		Family	Family + Territory		Family	Family + Territory
North	8.87	9.25		10.00	10.30		13.47	14.26
Centre	5.05	5.62		6.24	6.88		8.82	10.38
South and Islands	3.99	5.21		4.61	5.68		10.57	12.89
Italy	*6.39*	*7.20*		*7.37*	*8.30*		*11.96*	*16.13*

Notes: Family circumstances include gender, migration background, parental education, and parental occupational status. Territorial circumstances are proxied by province fixed effects. Values are percentages of variance explained ($R^2 \times 100$).

Table 2. Share of Variance in Reading Achievement Explained by Family and Territorial Circumstances, by Geographical Area and Educational Stage

	Grade 2			**Grade 5**			**Grade 8**	
Area	Family	Family + Territory		Family	Family + Territory		Family	Family + Territory
North	9.93	10.24		9.73	10.01		18.55	18.93
Centre	5.51	5.79		5.61	6.04		13.12	13.77
South and Islands	4.06	4.59		4.37	5.22		11.88	13.20
Italy	*6.98*	*7.40*		*6.95*	*8.46*		*15.05*	*16.88*

Notes: Family circumstances include gender, migration background, parental education, and parental occupational status. Territorial circumstances are proxied by province fixed effects. Values are percentages of variance explained ($R^2 \times 100$).

The longitudinal analysis reveals substantial persistence of educational achievement over time (Table 3). Prior achievement explains a larger share of subsequent outcomes in mathematics than in reading. However, family circumstances continue to exert an independent influence in both subjects, increasing the explained variance by 3.0 and 6.2 percentage points in mathematics and by 3.7 and 5.6 percentage points in reading. These findings suggest that educational inequalities continue to accumulate throughout compulsory schooling, particularly during the transition from primary to lower secondary education.

Table 3. Dynamic Contribution of Family Circumstances to Educational Achievement across Educational Transitions

	Prior achievement only (%)	**Prior achievement + family circumstances (%)**	**Additional variance explained by family circumstances (%)**
	Mathematics		
Grade 2 → Grade 5	28.75	31.79	3.04
Grade 5 → Grade 8	36.29	42.49	6.20
	Reading		
Grade 2 → Grade 5	18.92	22.65	3.73
Grade 5 → Grade 8	29.19	34.79	5.59

Notes: The first column reports the share of variance in educational achievement explained by prior achievement only. The second column reports the share of variance explained by prior achievement and observed family circumstances, including gender, migration background, parental education, and parental occupational status. Dynamic inequality of opportunity is measured as the additional share of variance explained by family circumstances after controlling for prior achievement.

School fixed–effects estimates indicate that most of the effects associated with parental education remain stable when students are compared within the same school. Preliminary interaction analyses also provide evidence of cumulative advantage, suggesting that students from more advantaged family backgrounds transform initial competencies into subsequent learning gains more effectively.

Conclusions

This paper contributes to the Equality of Opportunity literature by adopting a longitudinal perspective on educational inequality. Preliminary results suggest that observed family circumstances explain only a limited share of educational achievement, but their influence increases substantially throughout compulsory schooling. The analysis also reveals considerable geographical heterogeneity and provides evidence that educational inequalities continue to accumulate over time, particularly during the transition to lower secondary education. Future research will extend the analysis through a longitudinal application of Roemer's types–and–tranches approach.

Keywords: Educational Equality of Opportunity; Educational Inequality; Compulsory Schooling; Longitudinal Analysis

SESSION 11. FROM LATENT POTENTIAL TO OUTSTANDING ACHIEVEMENT: GIFTEDNESS, TOP PERFORMERS AND EXCELLENCE IN CONTEMPORARY EDUCATIONAL SYSTEMS

ORGANIZERS: UNIVERSITY OF MILANO BICOCCA – INVALSI
COORDINATORS: BRUNELLA FIORE – PAOLO BARABANTI
5TH NOVEMBER: 4.30 P.M. – 6.30 P.M. {AULA B0.1 – RESEARCH 15}

Fostering Collaborative Peer Relationships Among High–Potential and Gifted Children

Giorgia Ruzzante – Alessia Travaglini

Introduction

The promotion of social and prosocial skills represents an essential perspective for educational processes (Caprara&Bonino, 2006; Roche, 2002). As highlighted by numerous studies, prosociality constitutes a fundamental element of teaching that genuinely aspires to be inclusive (Travaglini, 2015; Cottini, 2021). In this regard, cooperative learning activities play a particularly important role because, when carefully designed, they enable students to experience forms of positive interdependence that simultaneously foster both cognitive and social development (Johnson et al., 2013).

Nevertheless, designing effective cooperative learning activities represents a complex challenge, as it requires teachers to develop open–ended tasks with multiple possible developments and solutions, which consequently require the equitable participation of all group members. This issue is especially relevant for high–potential and gifted learners. Depending on their individual characteristics—which should always be carefully considered—these children and adolescents may display heightened sensitivity to interpersonal conflicts with peers and family members (Neihart, 1999; Neihart, 2002), experience social isolation (Neihart, 1999), and report feelings of inadequacy and misunderstanding (Preuss & Dubow, 2004). An even more critical situation concerns students with so–called *twice exceptionality* (Milan, 2025), who combine giftedness/high potential with conditions such as ADHD or other comorbidities that are often difficult to identify.

Furthermore, the collaborative activities commonly proposed in schools tend to resemble traditional group work aimed merely at assessing knowledge according to a transmissive approach, rather than constituting a complex learning process requiring flexibility, critical thinking, and problem–solving skills. Consequently, when confronted with routine group tasks, high–potential or gifted students may experience a range of negative feelings, including:

- boredom;
- feeling "exploited" because they assume the role of mentor within the group, effectively replacing the teacher;
- perceiving their own effort as disproportionate to that of their classmates;
- experiencing a sense of unfairness, resulting in reluctance to participate in future group work;
- rejecting the final product because they perceive it as being of poor quality (Da Dalt & Da Re, pp. 103–104).

These observations highlight the need to rethink the ways in which cooperative learning activities are designed in schools. To fully embrace the principle of inclusion, such activities should enable all students to cooperate effectively toward the achievement of a common goal.

Research has demonstrated the effectiveness of educational strategies such as acceleration, enrichment, and curriculum compacting (Brazzolotto, 2019; Vinci & Sgambelluri, 2020; Pinnelli, 2026). Other promising perspectives include competency–based education (Castoldi, 2016) and authentic tasks (McTighe & Wiggins, 2014), interpreted through an inclusive educational framework (Travaglini, 2025).

Despite these advances, there remains a need for further research that more fully considers the perspectives of educational professionals (teachers and school leaders), as well as those of high–potential and gifted students and their families.

Specifically, the following research questions need to be addressed:

- How do teachers conceptualize and design cooperative learning activities?
- How is student heterogeneity taken into account in these activities?
- What difficulties do high–potential and gifted students experience during cooperative learning?
- What opportunities for improvement and educational intervention can be identified?

To address these questions, the study will adopt a mixed–methods approach based on the principle of data triangulation (Trinchero, 2002). On the one hand, teachers will be invited to describe their experiences in designing educational pathways that address classroom heterogeneity. On the other hand, the experiences and perceptions of high–potential and gifted students will be explored, also through the perspectives of their families.

Data Used

Building on the authors' exploratory study on giftedness (Ruzzante & Travaglini, 2023), the present research aims to collect qualitative data from both parents and teachers concerning their experiences with cooperative learning involving high–potential and gifted students.

Data will be gathered through questionnaires designed to explore participants' perceptions of the social and educational experiences of gifted learners within heterogeneous classrooms. Particular attention will be devoted to examining the social relationships that gifted students establish with their classmates, their participation in cooperative learning activities, the opportunities and barriers they encounter, and the strategies adopted by teachers to promote meaningful inclusion.

The study also seeks to investigate how parents perceive their children's experiences of collaboration with peers, including possible feelings of belonging, social isolation, underachievement, or frustration that may emerge during group work. Teachers' perspectives will complement these findings by providing insights into the planning, implementation, and management of cooperative learning activities, as well as the challenges associated with addressing classroom heterogeneity and responding to the educational needs of gifted learners.

The integration of these complementary perspectives will enable methodological triangulation (Trinchero, 2002), providing a more comprehensive understanding of the conditions that facilitate or hinder inclusive cooperative learning for high–potential and gifted students. The resulting dataset will support the identification of effective educational practices and inform the development of inclusive teaching strategies that promote both academic achievement and social participation.

Methodology

Data collection will take place at the beginning of the 2026/2027 scholastic year (September 2026). Participants will be recruited through convenience sampling and will include teachers from all educational levels—from preschool and primary education to lower and upper secondary schools—working in different regions of Italy. To ensure a comprehensive understanding of the phenomenon, the study will also involve parents of high–potential and gifted students, whose perspectives are considered essential for interpreting students' educational and social experiences beyond the school context.

Data will be collected primarily through structured questionnaires comprising both closed–ended and open–ended questions. The questionnaires will investigate teachers' approaches to planning and implementing cooperative learning activities, the strategies adopted to address classroom

heterogeneity, the perceived participation of gifted students in collaborative tasks, and the main challenges encountered in promoting inclusive educational practices. Parents will be invited to report their perceptions of their children's social relationships with peers, their experiences of cooperative learning, and any difficulties or positive outcomes observed in the school context.

Participants will be recruited through online dissemination of the questionnaires via social media platforms, particularly through professional networks of teachers and online communities of parents of high–potential and gifted students. This recruitment strategy is expected to facilitate access to participants from different geographical areas and educational settings, thereby increasing the diversity of the sample.

Data analysis will follow a mixed–methods research design, integrating quantitative and qualitative approaches. Quantitative data will be analysed using descriptive statistical techniques to examine the distribution of responses, identify recurring patterns, and compare results across different educational levels. Qualitative responses to the open–ended questions will undergo thematic analysis in order to identify the main themes, critical issues, and examples of effective educational practices.

The integration of quantitative and qualitative findings will enable methodological triangulation, providing a richer and more comprehensive understanding of how cooperative learning is conceptualised and implemented for high–potential and gifted students. Particular attention will be devoted to identifying inclusive teaching strategies and good practices that may inform future teacher education programmes and support schools in designing learning environments that foster both academic development and social participation for gifted learners.

Results and discussion

The study aims to contribute to advancing knowledge and awareness of giftedness within both initial teacher education and continuing professional development programmes. Despite the growing international interest in gifted education, this topic is still addressed inconsistently across Italian universities and regional educational authorities. As a result, many teachers enter the profession with limited preparation for recognising the characteristics of giftedness and designing appropriate educational responses, while opportunities for in–service training remain uneven across the country.

The findings are expected to provide good practices on how cooperative learning is currently conceptualised and implemented in heterogeneous classrooms, with particular attention to the experiences of high–potential and gifted students. By comparing the perspectives of teachers and parents, the study will identify the strengths and limitations of existing educational practices, highlighting the factors that facilitate or hinder the meaningful participation of gifted learners in collaborative activities.

A further expected outcome is the identification of examples of effective inclusive practices that can support teachers in designing cooperative learning environments capable of responding to the educational needs of all students without compromising the learning opportunities of those with high cognitive potential. In particular, the study seeks to promote educational strategies that foster positive peer relationships, prosocial behaviours, engagement, and academic challenge while preventing boredom, underachievement, disengagement, and social isolation, which are frequently reported among gifted learners.

More broadly, the research is expected to contribute to the ongoing debate on inclusive education by emphasising that giftedness should be recognised as a form of neurodiversity requiring specific educational attention. Just as schools develop personalised educational pathways for students with disabilities, specific learning disorders, or other special educational needs, they should also provide differentiated and appropriately challenging learning opportunities for gifted and high–potential students. Such an approach is fully consistent with the principles of inclusive education, which aim to ensure that every learner can develop his or her potential within a supportive and equitable educational environment.

Finally, the study intends to offer practical recommendations for teacher education, school leadership, and educational policy. The results may inform the design of teacher training

programmes, support schools in developing good inclusive practices, and encourage greater institutional recognition of gifted education within the Italian educational system. In doing so, the research seeks to promote a more comprehensive understanding of inclusion—one that embraces the full spectrum of learner diversity and recognises giftedness as an educational need deserving appropriate pedagogical responses.

Conclusions

The proposed study is of particular relevance because this topic remains largely underexplored within the Italian educational context. While research on giftedness has increasingly focused on the identification of high–potential learners and on instructional strategies aimed at promoting academic achievement, considerably less attention has been devoted to the social dimension of learning and to the ways in which gifted students experience collaboration within heterogeneous classrooms.

In the Italian school system, discussions of gifted education have traditionally centred on cognitive characteristics and individualised educational interventions. By contrast, relatively little is known about how gifted students participate in cooperative learning activities, how they perceive interactions with their peers, or how classroom relationships influence their engagement, motivation, and sense of belonging. This represents a significant gap, particularly in light of the growing emphasis on inclusive education, which recognises learning as a fundamentally social process and values participation, collaboration, and mutual support among students with diverse educational needs.

The social dimension of learning is especially important for high–potential and gifted students because their educational experience extends beyond academic performance. Previous research has shown that many gifted learners may experience difficulties in peer relationships, feelings of isolation, perfectionism, heightened sensitivity, or frustration when collaborative activities fail to provide appropriate intellectual challenge or equitable participation. At the same time, well–designed cooperative learning environments can foster prosocial behaviours, empathy, communication skills, and reciprocal learning, while creating opportunities for gifted students to contribute meaningfully without assuming inappropriate leadership or tutoring roles.

By investigating the perspectives of both teachers and parents, the present study seeks to provide a comprehensive understanding of how cooperative learning is currently implemented and experienced in Italian schools. This dual perspective will make it possible to identify convergences and discrepancies between educational practices and family perceptions, offering valuable insights into the factors that facilitate or hinder the inclusion of gifted learners within collaborative settings.

Ultimately, the study aims to contribute not only to the national debate on gifted education but also to the broader international discussion on inclusive pedagogy. By highlighting the importance of the social dimension of learning, the findings are expected to support the development of evidence–based educational practices that promote both academic excellence and socio–emotional well–being. In doing so, the research advocates for a more comprehensive conception of inclusion—one that recognises giftedness as a form of learner diversity requiring educational responses that are academically challenging, socially meaningful, and responsive to the individual characteristics of every student.

Keywords: Giftedness; High–potential students; Inclusive education; Cooperative learning

Improving the Identification of Gifted and Talented Students in the Basque Country to Ensure Educational Inclusion and Full Development

Miryam Martínez-Izaguirre – Alexander Álvarez-González – Arantza Arruti – Ioseba Iraurgi

Introduction

Current inclusive education demands a transformation in how school systems identify and address diversity, moving beyond traditional approaches focused solely on deficit compensation. Historically, the lack of systematic attention has rendered the specific educational support needs of gifted and talented students invisible, perpetuating the false belief that high intellectual potential guarantees academic success on its own. Recent scientific research refutes this myth, demonstrating the importance of accompanying and promoting talent development (Gagné, 2015; Olszewski–Kubilius et al., 2015) and highlighting the vulnerability of these students (Martínez–Izaguirre et al., 2024). Without proper intervention, the crystallization of talent fails, exposing students to underachievement, academic disengagement, or school failure (Barbier et al., 2019; Barrera–Algarín et al., 2023; McCoach et al., 2020).

Acknowledging this reality, the Basque educational system aims to transform it. To achieve this, the key pillars established are the importance of raising awareness and training educational professionals, and identifying these students to determine the specific areas where they require educational support.

Data Used

The analysis and justification of inclusion policies in the Basque Country are supported by evidence gathered in a prior study conducted by the Basque Institute for Educational Research and Evaluation (ISEI–IVEI) in 2019. The data from this preliminary study reflected an alarming situation of low identification in the Basque Country, with barely 747 officially identified students, representing a prevalence rate of only 0.2% of the total school population (Basque Government, 2019). Parallelly, this baseline study unveiled a deep contradiction in school culture: although schools perceived widespread training needs, there was a notable lack of priority from school management teams toward training and addressing gifted education, given that over 65% of school principals stated that this matter was not a priority within their school's strategic management plan.

Once the needs were analyzed, the plan designed and implemented by the University of Deusto and the Department of Education of the Basque Government generated the following main changes. It is mandatory to establish measures that foster environmental and personal catalysts to transform potential into expressed talent. To this end, the proactive and systematic search for talent becomes a fundamental guarantee of equity and human rights (Antunes et al., 2022; Ninkov, 2020), ensuring that disadvantaged socioeconomic conditions or gender biases do not act as barriers that conceal student potential. For this purpose, two main actions are necessary: having qualified, aware, and trained professionals to identify potential and promote its development, and having tools that guide such processes. Thus, the combination of both measures transforms the school reality as follows:

Currently, the Department of Education reports a total of 7,237 gifted and talented students in the Basque Country. This figure comprises 4,513 students identified via deliberate universal screening and 2,724 via traditional incidental pathways.

Methodology

The methodology applied to transform the Basque educational reality is structured around two inseparable pillars: a systematic teacher training plan and the design, validation, and application of a screening instrument.

Systematic Teacher Training Plan

In parallel with the technical deployment of the SACI, systematic training has been developed to empower different agents within the educational system. This compulsory and continuous training has specifically addressed the following essential aspects and contents:

- Awareness and Demolishing Myths. Content aimed at eradicating social and educational stereotypes that associate giftedness with guaranteed success or homogeneous behavioral profiles, particularly protecting vulnerable groups (girls or disadvantaged backgrounds).
- Early Detection and Identification of Signs. Practical training in recognizing behavioral, cognitive, and socio–emotional indicators of giftedness from Early Childhood and Primary Education.
- Inclusive Educational Intervention and Response. Applied training in designing personalized plans, multi–level programming strategies, curricular enrichment methodologies within the ordinary classroom, and regulatory criteria for grade acceleration.
- Unification of Technical Criteria. Specific training for psycho–pedagogical advisory services to homogenize action guidelines and counseling for schools.
- Initial Training. Systematic inclusion of modules on educational attention to gifted and talented students in degrees in Education.

Validation and Implementation of the SACI Screening Protocol

To overcome the biases of incidental detection (which is reactive and linked to stereotypes), the screening instrument (SACI) was designed and validated. In its validation, the sensitivity of the instrument was intentionally prioritized over its specificity, maximizing identification capacity to prevent any student with high potential from being excluded (assuming potential false positives that are refined in subsequent phases).

This validated protocol has already been implemented systematically and compulsorily in all schools in the Basque Country at strategic developmental milestones, specifically in the 1st and 6th grades of Primary Education.

Results and Discussion

The joint implementation of systematic training and the SACI Protocol has triggered a profound transformation of the school reality in the Basque Country, visible both in teacher qualification and in quantitative identification results.

Transformation of Identification and Clear Data

Current data demonstrate a historic breakthrough in terms of making this student group visible. The system has grown from 747 cases in 2019 to a cumulative total of 7,237 identified students.

Deliberate universal screening accounts for most of this transformation with 4,513 cases, proving the effectiveness of actively seeking talent instead of waiting passively, while the incidental pathway contributes 2,724 cases.

Regarding specific identification per academic year in which the screening is applied, the data remain stable within a range of 3.9% to 4.2% of the enrolled student body. Although this percentage does not reach the maximum theoretical estimate in scientific literature (which places high potential between 10% and 15%), it represents a radical break from the initial 0.2%, far exceeding the national average of 0.7%.

Value Placed by Teachers on Received Training

The reach of systematic training expanded significantly through the use of the streaming modality starting in the 2023–2024 academic year. Participation data record a massive evolution: 3,765 connections in the 2023–2024 course, 4,312 connections in the 2024–2025 course, and 4,100 direct participating teachers in the 2025–2026 course, accumulating a minimum total of 12,177 trained teachers in the Basque Country. Teachers have highly valued this training, as it equips professionals with real tools to respond in a personalized manner to students in their ordinary classrooms and transforms their technical perception.

Discussion

The discussion of these results evidences that the historical underidentification did not stem from a real absence of talent in the Basque Country, but rather from the structural deficiencies of an incidental detection model that was blind to specific educational support needs. By introducing a validated instrument (SACI), the system has effectively neutralized demographic and structural

variables, given that differences in detection rates between linguistic models are below 1% and the maximum territorial variability is barely 0.6% in Primary Education and 1.5% in Compulsory Secondary Education.

Nonetheless, the discussion must emphasize overcoming the initial resistance from school management teams. Massive and applied training has been the indispensable catalyst to resolve the 2019 contradiction. By equipping teachers with intervention and enrichment competencies,giftedness is no longer perceived as an administrative burden or an elite myth, becoming integrated as a real regulatory priority instead. Likewise, the stabilization of rates per course (3.9%–4.2%) opens the debate on the need to keep refining specific grading and norming criteria to correct the gender gaps that still persist (where identification in boys remains higher, 4.6% vs. 2.9% in Primary Education), proving that the inclusion process is continuous and requires keeping both active training and a sensitive teacher perspective alive above mere labeling.

Conclusions

The process of implementing the new inclusive model in the Basque Country allows for drawing fundamental conclusions:

The prior study from 2019 proved that reactive and incidental approaches condemn talent to invisibility (0.2% identification) due to a lack of strategic priority in schools.

The validation and systematic application of the universal SACI protocol at key milestones (1st grade of Primary and the transition to Secondary Education) constitutes an indispensable tool to guarantee equity and social justice in the educational system.

Systematic training of more than 12,000 teachers in key aspects (awareness, detection signs, and classroom intervention) is the only mechanism capable of giving real value to identification, transforming technical instruments into personalized and effective educational responses.

Keywords: Gifted students; Talent; Identification; Compulsory education

Profiles and Pathways of Learning Strategy Development During the Transition to Higher Education: Evidence from High Ability and Typical Students

Alicia Ramos – Vincent Donche – Karine Verschueren

Introduction

The transition from secondary to higher education represents a critical period in students' academic development, characterized by substantial shifts in instructional practices, learning expectations, and student autonomy. This transition has been described as a "learning shock," as students must adapt to reduced instructional guidance, fewer evaluation moments, and increased demands for self–directed learning (Christie, 2009). In this context, students are expected to take responsibility for planning, monitoring, and regulating their own learning processes, often with limited external support. Successfully navigating this transition requires the flexible deployment of a well–developed repertoire of learning strategies.

Learning strategies consist of both cognitive processing strategies and regulation strategies. Cognitive processing strategies range from the use of surface learning activities, such as memorization aimed at reproducing information, to deep learning activities involving the integration and structuring of knowledge. Regulation strategies include self–regulation—where students actively guide and evaluate their learning—and lack of regulation, which reflects aimless and poorly structured learning behavior (Vermunt & Donche, 2017; Zeidner, 2019). These strategies develop dynamically over time and are shaped by ongoing interactions between individual characteristics

and contextual factors such as instructional practices and motivational beliefs (Coertjens et al., 2017; Donche et al., 2013; Kaplan et al., 2020).

For high–ability students, research suggests that many of them have failed to develop an adequate repertoire of learning strategies before entering higher education. Although these students possess strong cognitive abilities, qualitative research suggests that they may not consistently develop adaptive learning strategies during compulsory education due to insufficient curricular challenge (Ramos et al., 2022). When learning tasks do not require sustained effort or strategic engagement, students may rely on intuitive or "effortless" approaches to learning, limiting opportunities to develop self–regulatory skills. Consequently, when these students encounter the increased demands of higher education, they may experience difficulties in adapting their learning behavior. Evidence from the Flemish context indicates that a substantial proportion of high–ability students experience study delay in higher education at rates comparable to their typical peers (Ramos et al., 2019).

Motivational factors are central to understanding these processes. According to expectancy–value theory, students' engagement depends on their beliefs about their competence (self–efficacy) and the value they assign to learning tasks, including intrinsic, attainment, and utility value (Wigfield & Eccles, 2000). The Achievement Orientation Model further posits that effective learning behavior depends on the interaction between ability, motivation, and environmental support (Siegle et al., 2017). Perceived curricular challenge also plays a crucial role in the development of both motivation and learning development, as appropriately challenging tasks foster both motivational engagement and the development of effective learning strategies (Dignath & Veenman, 2021; Lavrijsen et al., 2021). Appropriate curricular challenge is important for all students but more difficult to provide to high–ability students within general classrooms.

Despite the theoretical importance of these relationships, longitudinal, person–oriented research examining learning strategy development across the transition to higher education—particularly among high–ability students—remains limited (Shum et al., 2024). This study addresses this gap by examining how learning strategy profiles emerge and evolve across this transition and how they are shaped by cognitive ability, motivation, and perceived curricular challenge. Specifically, the study investigates the following research objectives (ROs): identification and stability of learning strategy profiles (RO1 and RO2), differences between high–ability and typical students (RO3), the role of motivational and contextual antecedents (RO4), and the implications of these profiles for academic adjustment and intellectual development in higher education (RO5).

Data Used

The study uses longitudinal data from the TALENT project, a large–scale study conducted in Flanders. The original sample consisted of 3,409 students followed from the beginning of secondary education. For the present analyses, data from the final year of secondary education (n = 870) and the first year of higher education (n = 589) were used.

Cognitive ability was assessed at the start of secondary education using a subtest measuring fluid intelligence from the CoVaT–CHC instrument, a standardized intelligence test developed for and validated in the Flemish context (Magez, 2015). High–ability students were operationalized as those scoring within the top 10% of scores from a representative norm group, consistent with established definitions of giftedness (Gagné, 2004). Students whose scores fell within one standard deviation of the mean were considered to have average ability, and were used as a comparison group.

Learning strategies were measured using the Learning and Motivation Questionnaire (LEMO), which captures, among other constructs, the constructs memorizing, relating and structuring, self–regulation, and lack of regulation that were used for the current study (Donche et al., 2010). Motivational variables included intrinsic value, attainment value, utility value, self–efficacy, goal commitment, and amotivation, assessed using validated instruments (Conley, 2012; Klein et al., 2001). Perceived curricular challenge was measured at the end of secondary education (Lavrijsen et al., 2021). Outcomes in higher education included academic adjustment and intellectual

development, measured using established scales (Baker & Siryk, 1986; Pascarella & Terenzini, 1980). All measures show acceptable reliability and have been validated in the Flemish context.

Methodology

A person–oriented analytical approach was adopted to capture heterogeneity in students' learning strategy use. Latent Profile Analysis (LPA) was conducted separately at the end of secondary education and the beginning of higher education to identify distinct learning strategy profiles based on the four LEMO indicators. Model selection was guided by statistical fit indices, classification accuracy, parsimony, and interpretability.

To examine developmental changes, Latent Transition Analysis (LTA) was conducted to estimate stability and transitions between profiles across the two educational stages (Muthén & Asparouhov, 2022). Cognitive ability was included as a predictor of profile membership and transitions using the BCH method, an auxiliary modeling approach (Asparouhov & Muthén, 2021). The BCH method was also used to examine motivational variables and perceived curricular challenge as antecedents of profile membership and transitions. Finally, the predictive validity of the profiles was assessed by linking them to academic adjustment and intellectual development in higher education, also by using the BCH method.

Results and Discussion

The latent profile analyses identified a three–profile solution at both time points. The first profile, Strategic and Self–Regulated, was characterized by high levels of deep processing and self–regulation. The second profile, Moderate Processing and Self–Regulation, reflected moderate use of both cognitive and regulatory strategies. The third profile, Low Processing and Self–Regulation, was characterized by low levels across these dimensions. These profiles align with theoretical distinctions between adaptive and maladaptive learning patterns (Vermunt & Donche, 2017).

The transition analysis indicated that learning strategy profiles were relatively stable across the transition to higher education. Most students remained in the same profile, and transitions occurred primarily between adjacent profiles. No direct transitions were observed between the most and least adaptive profiles, suggesting that changes in learning behavior tend to be gradual rather than abrupt. Importantly, high–ability students were significantly less likely to belong to the Strategic and Self–Regulated profile compared to their peers. This finding suggests that high cognitive ability does not necessarily translate into the use of adaptive learning strategies and may, under certain conditions, even be associated with less developed strategic learning behavior..

Consistent with theoretical expectations, perceived curricular challenge emerged as a strong predictor of adaptive learning strategy profiles. Students who experienced higher levels of challenge were more likely to belong to both the moderate and strategic/self–regulated profiles. In line with this pattern, high–ability students also reported lower levels of perceived curricular challenge during secondary education. These findings provide empirical support for the hypothesis that insufficient challenge in earlier educational contexts may limit opportunities for high–ability students to develop effective learning strategies (Ramos et al., 2022). When students are not required to engage in effortful and strategic learning, they may not acquire the skills necessary to regulate their learning independently. As a result, they may be less prepared for the demands of higher education, despite their cognitive potential.

Motivational variables also played a significant role in predicting learning behaviors, with intrinsic value and attainment value showing particularly strong associations with adaptive profiles. These findings align with expectancy–value theory, which emphasizes the importance of task value in motivating engagement with learning strategies (Wigfield & Eccles, 2000). Furthermore, it is congruent with previous learning pattern research (i.e. Vermunt & Donche, 2017), in which adaptive learning strategies were found to relate to intrinsic values.

Finally, the results demonstrated that learning strategy profiles are meaningfully related to outcomes in higher education. Students in the Strategic and Self–Regulated profile reported higher levels of academic adjustment and intellectual development than their peers in less adaptive profiles. This

finding confirms the importance of self–regulated and deep learning strategies for successful adaptation to the higher education environment.

Conclusions

This study provides a longitudinal, person–oriented examination of learning strategy development across the transition to higher education, with a particular focus on high–ability students. The findings demonstrate that learning strategies cluster into distinct profiles that are relatively stable over time in the transition to higher education and strongly related to academic outcomes.

A central contribution of this study is the finding that high–ability students are less likely to be represented in the most adaptive learning strategy profile and report lower levels of perceived curricular challenge during secondary education. Together with the lower levels of perceived curricular challenge during secondary school reported by these students, these results suggest that underchallenge may play a key role in shaping the developmental trajectory of learning strategies among high–ability students. In the absence of sufficiently demanding learning environments, these students may not develop the self–regulatory and deep processing strategies required for success in higher education.

The findings highlight the importance of considering the interaction between ability and educational context. Cognitive ability alone is insufficient to ensure adaptive learning behavior; rather, the development of effective learning strategies depends on opportunities to engage with challenging tasks and on the presence of supportive motivational beliefs.

From a practical perspective, these results underscore the need for educational practices that provide adequate challenge and explicitly support the development of learning strategies, particularly for high–ability students. Interventions aimed at improving the transition to higher education should focus not only on academic content but also on fostering self–regulation, motivation, and strategic learning behavior.

Keywords: Learning strategies; High ability students; Transition to higher education; Person–centered analysis

Teachers' attitudes toward gifted education in Italy: validation of the ATGE scale and a mediation model linking attitudes, meritocratic beliefs, and endorsement of school selection

Francesco Marsili – Maria Miranda – Annalisa Morganti

Background and rationale

Contemporary education systems face a fundamental tension between inclusive ideals and selective practices. While international frameworks assert all learners' right to meaningful participation and high–quality education (UN, 2006; UNESCO, 1994), schooling remains organized around competitive, meritocratic mechanisms that rank and reward performance (Batruch et al., 2019). Gifted education offers a revealing lens for examining these dynamics: contemporary models emphasize dynamic, contextual conceptions of giftedness (Subotnik et al., 2023), yet gifted education has historically functioned as a selection mechanism that reproduces hierarchies (Borland, 2005).

Teachers' attitudes are a key element of this picture, as they can act as barriers or facilitators of gifted students' development and shape the implementation of inclusive and differentiated practices. A recent systematic review showed that, although teachers generally express positive attitudes toward meeting gifted students' needs, they are ambivalent toward interventions that remove students from regular classrooms (Sambuis & Bourdin, 2024). At the ideological level, belief in school meritocracy (BSM)—the conviction that academic success reflects individual effort and ability—operates as a lens that legitimizes structural inequalities and reduces support for equity–oriented policies (Batruch et

al., 2023; Darnon et al., 2018; Wiederkehr et al., 2015), while belief in the selective function of schooling (BSS) refers to the conviction that education's central role is sorting and rewarding high performers through competitive mechanisms (Autin et al., 2015).

In Italy—a context characterized by a strong inclusive tradition but an increasing policy emphasis on merit and excellence, with recent regulatory initiatives concerning gifted learners—research on teachers' attitudes toward gifted education has been scarce, and no validated instrument was available to measure them. This contribution presents two sequential studies from the same research program: Study 1 provides the Italian validation of the Attitudes Towards Gifted Education (ATGE) scale (Laine et al., 2019); Study 2 uses the validated instrument to test a mediation model linking attitudes toward gifted education, meritocratic beliefs, and endorsement of the selective function of schooling.

2. Study 1: Italian validation of the ATGE scale

2.1 Aims and method

Study 1 aimed to investigate the factor structure, reliability, and validity of the Italian adaptation of the ATGE scale, originally developed for the Finnish context (Laine et al., 2019) and rooted in the Opinions About the Gifted and Their Education questionnaire (Gagné & Nadeau, 1985). The sample included 513 Italian in–service teachers (82.8% female; Mage = 38.53), both curricular and support teachers, from all school levels. The scale covers eight dimensions: special support and social value, elitism, objections to support, support for differentiation, practical obstacles to serving gifted learners, support for gifted acceleration, prioritization of special education for children with difficulties, and educational challenges for gifted children. A confirmatory factor analysis (CFA) with WLSMV estimation was performed, and convergent validity was examined against measures of attitudes toward inclusive education and perceived teaching competencies.

2.2 Results

The CFA supported the eight–factor solution of the ATGE, $\chi^2(349) = 1195.80$, $p < .001$, CFI = .91, TLI = .90, RMSEA = .06. Seven items with factor loadings below .30 were removed, yielding a 29–item final version with acceptable internal consistency (Cronbach's $\alpha = .76$). Analyses of group differences showed that support teachers reported higher levels on the elitism dimension than curricular teachers. Overall, the results indicate that the Italian ATGE is a psychometrically robust instrument for assessing teachers' attitudes toward gifted education, filling a measurement gap in the Italian context.

3. Study 2: Attitudes, meritocratic beliefs, and school selection

3.1 Aims and hypotheses

Study 2 examined how teachers' attitudes toward gifted education (ATGE) relate to belief in school meritocracy (BSM) and endorsement of the selective function of schooling (BSS). We tested a mediation model in which ATGE predicts BSS both directly and indirectly through BSM, hypothesizing that (H1) favorable attitudes toward gifted education are positively associated with endorsement of school selection; (H2) this relationship is mediated by belief in school meritocracy; and (H3) the combined direct and indirect effects indicate that the more teachers support gifted education, the more they perceive schooling as a meritocratic and selective system.

3.2 Method

Participants were 463 Italian in–service teachers (60.5% curricular, 39.5% support; 65.0% aged 41–60; 53.6% with more than ten years of experience). Although most had attended training on inclusion (73.4%) and disability (73.7%), only 10.4% had received training specifically focused on gifted education. Teachers completed the Italian ATGE validated in Study 1 ($\alpha = .76$), the Italian adaptation of the Belief in School Meritocracy scale (Wiederkehr et al., 2015), and the Italian adaptation of the scale measuring the selective function of schooling (Autin et al., 2015). The mediation model was tested with teaching role included as a covariate.

3.3 Results

No significant differences emerged between curricular and support teachers in ATGE ($t = 0.48$, $p = .633$, $d = 0.05$); however, curricular teachers reported significantly higher BSM ($t = 2.93$, $p = .004$, $d = 0.28$) and BSS ($t = 2.16$, $p = .031$, $d = 0.20$). ATGE correlated positively with BSM ($r = .10$, $p < .05$) and BSS ($r = .26$, $p < .01$), and BSM correlated with BSS ($r = .32$, $p < .01$). In the mediation model, ATGE showed a significant positive direct effect on BSS ($B = 0.57$, $SE = .11$, $p < .001$) and on BSM ($B = 0.16$, $SE = .07$, $p = .030$); BSM significantly predicted BSS ($B = 0.45$, $SE = .07$, $p < .001$); and the indirect effect of ATGE on BSS through BSM was significant ($B = 0.07$, $SE = .03$, $p = .039$), indicating partial mediation. Teaching role was significantly associated with BSM, and its indirect effect on BSS through BSM was also significant.

4. General discussion and implications

Taken together, the two studies deliver both a methodological and a substantive contribution. Methodologically, they provide the Italian research community with a validated instrument for assessing teachers' attitudes toward gifted education. Substantively, they show that favorable attitudes toward gifted education are associated with stronger endorsement of meritocratic beliefs and of the school's selective function, with meritocratic beliefs acting as a partial mediator. While support for gifted education is widely shared across professional roles, ideological beliefs diverge systematically between curricular and support teachers, reflecting their differential positioning within evaluative and equity–oriented institutional functions.

The findings illuminate a fundamental asymmetry in how meritocratic ideology operates across student populations: the same belief system that facilitates support for gifted education has been shown to undermine support for the inclusion of students with disabilities (Khamzina et al., 2021; Stanczak et al., 2024). Invoking meritocratic principles may thus mobilize support for gifted education while reinforcing hierarchical views that privilege excellence over equity. Teacher education should address these ideological dimensions directly, strengthening teachers' self–efficacy in within–classroom differentiation and promoting framings that emphasize diversity and collective responsibility, so as to reconcile excellence with inclusion (Elias et al., 2025; Sambuis & Bourdin, 2024). Limitations include the cross–sectional design, which precludes causal inference, the reliance on self–report measures, and the modest magnitude of the indirect effect, which points to additional mechanisms—such as self–efficacy, personal interactions with gifted learners, and essentialist beliefs about ability—to be investigated in future research.

Keywords: Gifted education; Teacher attitudes; Scale validation; School meritocracy; School selection; Inclusive education

The state of intervention research in gifted education: a scoping review of experimental and quasi–experimental studies (2010–2025)

Francesco Marsili – Marta Pellegrini – Patricia Ayllón Salas – Luigi Fenza

Background and rationale

Over the past three decades, research in gifted education has been repeatedly scrutinized for its methodological robustness. Early systematic analyses described a fragmented scientific community characterized by recurring methodological weaknesses, including the predominance of observational quantitative measures and the limited use of rigorous experimental designs and control groups (Ziegler & Raul, 2000). Subsequent reviews confirmed structural shortcomings such as the absence of reliable pre–post measures and controlled designs, which constrain the possibility of drawing robust causal inferences (Bailey et al., 2008). A large–scale analysis of more than one thousand studies found that empirical work in the field remained largely descriptive or correlational, with fewer than 10% of studies employing true experimental designs (Dai et al., 2011). More recently, a systematic review of educational interventions with gifted students found that none of the included studies used randomization or an active control group (García–Martínez et al., 2021).

Second–level analyses and scoping reviews have broadened the thematic perspective of the field while shedding further light on its methodological landscape. A structural topic model of over 5,000 articles across nine leading gifted education journals showed that equity and policies have become the most studied topic, while curriculum and instruction—historically dominant—have declined in relative intensity (Şakar & Tan, 2025). A recent scoping review found that, although methodological standards have improved over time, the field continues to grapple with heterogeneous definitions of giftedness and a restricted thematic scope that leaves entire areas without robust meta–analytic coverage (Popp et al., 2025). An overview of meta–analyses on giftedness similarly showed that most syntheses drew on comparative study designs, with only a minority requiring experimental or quasi–experimental studies as an inclusion criterion (Ogurlu, 2020). A scoping review of differentiated instruction practices for high–ability learners likewise drew on a methodologically heterogeneous body of evidence, with no studies excluded on design grounds (Nicholas et al., 2024), while a rare exception explicitly incorporating quasi–experimental designs was limited to a final sample of ten studies (Elballah et al., 2024).

Taken together, these reviews suggest that rigorously controlled designs remain underrepresented in gifted education, limiting the causal inferences that can be drawn from the existing literature. Yet, to date, no review has systematically and comparatively analyzed the use of experimental and quasi–experimental designs in gifted education, nor thoroughly examined the methodological characteristics of such studies in the most recent research landscape. Considering the theoretical and methodological transformations that have affected the field over the past fifteen years—including greater attention to equity, inclusion, and evidence–based evaluation—a new scoping review focused on studies published between 2010 and 2026 is warranted.

Objectives

This scoping review aims to map and appraise the state of causal research on the effectiveness of school–based educational interventions for gifted students (K–12). It is guided by three primary objectives:

(1) to map the types of interventions, outcomes, and student population characteristics that have been studied, in order to identify areas with a substantial evidence base and areas where important gaps remain;

(2) to examine the methodological approaches used (e.g., RCTs, quasi–experimental designs, single–case designs, pre–post designs) to determine the credibility and strength of the empirical evidence; and

(3) to assess how study results are reported, with particular attention to effect size reporting, as an indicator of the field's transparency and quantitative maturity.

Method

Protocol and registration

The review protocol was preregistered on the Open Science Framework and is publicly available at https://osf.io/t7vdr/. The review is conducted in accordance with established guidance for scoping reviews and systematic evidence syntheses (Page et al., 2021; Tricco et al., 2018). Study selection and data extraction are managed in Covidence.

Eligibility criteria

Table 1 summarizes the eligibility criteria, organized according to the PICOS framework.

Table 1. Eligibility criteria

Criterion	Specification
Population	Students identified as gifted or talented attending K–12. Studies including gifted students and their classmates are eligible only if a separate analysis for gifted students is reported.
Intervention	School–based strategies, interventions, or programs, delivered by teachers, specialists, or peers, within regular curricular instruction or school–supervised extracurricular activities. Both gifted–specific approaches (e.g., acceleration, enrichment, differentiation, curriculum compacting, mentoring, ability grouping) and non–specific strategies applied to gifted learners (e.g., cooperative learning, peer tutoring, cognitive strategy instruction, UDL) are eligible. Exclusively out–of–school or summer programs are excluded.
Outcomes	All quantitatively measured student outcomes: academic and cognitive (e.g., achievement, executive functions, thinking skills) and non–cognitive (e.g., self–regulation, motivation, socioemotional skills, wellbeing, creativity). Outcome domain is a descriptive variable, not an inclusion filter.
Study designs	(1) Randomized controlled trials; (2) quasi–experimental designs with a comparison group (matched/non–equivalent control group designs, difference–in–differences, regression discontinuity, instrumental variable analyses); (3) single–group designs (pre–post comparisons, interrupted time series); (4) single–case experimental designs.
Other criteria	Studies published between 2010 and 2025; no language or publication–type restrictions; grey literature (dissertations, reports, working papers) eligible; conference abstracts and proceedings excluded.

Search strategy

Systematic searches were conducted in February 2026 in six databases: ERIC (ProQuest), Social Sciences Citation Index (Web of Science), Scopus, APA PsycInfo and PsycArticles (ProQuest), Education Source (EBSCOhost), and ProQuest Dissertations and Theses Global. The search strategy combined three blocks of keywords related to the target population, the school setting, and experimental and quasi–experimental research designs. Given the lack of terminological consensus and the historical evolution of the concept of giftedness, a broad and inclusive set of population–related terms was used to maximize sensitivity (e.g., gifted, talented, high ability, highly able, highly capable), together with database–specific controlled vocabulary. The search string was validated against a set of known seed articles and adapted to each database. Complementary strategies include

hand–searching six specialist journals (e.g., Gifted Child Quarterly, Journal for the Education of the Gifted, Roeper Review) and the websites of major organizations in the field (e.g., NCRGE, Renzulli Center, Belin–Blank Center, ECHA, NAGC, World Council for Gifted and Talented Children), as well as backward and forward citation searches of previous reviews.

Screening

Study selection follows a two–stage procedure. At the title–and–abstract stage, two reviewers independently screened records using ASReview, an active learning–based screening tool, following a four–phase workflow adapted from the SAFE procedure (Screen, Active learning, Find more, Evaluate; Boetje & van de Schoot, 2024). This procedure combines multiple conservative stopping criteria (minimum coverage thresholds, retrieval of all seed articles, and runs of consecutive irrelevant records), and includes a model–switching phase with a deep semantic model to identify hard–to–find records, as well as an independent quality–control phase conducted by a third reviewer to minimize false negatives. Records retained at this stage are assessed in full text in Covidence by two independent reviewers against the eligibility criteria; inter–rater agreement is quantified with Cohen's kappa, and disagreements are resolved by a third researcher.

Data extraction

Data are extracted in Covidence using a structured, pre–specified codebook developed according to the PICOS framework and pre–tested on seed studies. For all included studies, core characteristics are coded: bibliographic information, population characteristics (country, grade level, gifted identification method), intervention and control characteristics (type, duration, intensity, nature of the control condition), outcome domains, study design, and sample sizes. For RCTs and quasi–experimental designs, a supplementary set of methodological variables is extracted to appraise internal validity, external validity, and reporting quality, including baseline equivalence, level of assignment, implementation fidelity, attrition level and handling, overall rating against What Works Clearinghouse standards (WWC, 2022), outcome measure quality, effect size type and completeness of reporting, adjustment for covariates, estimand type, and sufficiency of data for potential meta–analytic synthesis. Extraction is performed by one coder and validated by a second experienced coder, with disagreements resolved in regular team meetings.

Synthesis

Consistent with the scoping review methodology, findings will be synthesized narratively and in tabular and graphical form, mapping the distribution of interventions, outcomes, populations, and study designs, and profiling the methodological quality and reporting practices of the group–design studies.

Current status and preliminary work

At the time of writing, the database searches, the title–and–abstract screening, and the full–text eligibility assessment workflow have been completed or are in their final stages. Several hundred records retained through the SAFE–based screening have been advanced to full–text review in Covidence, and a substantial set of studies meeting the eligibility criteria has already been identified, spanning the full 2010–2026 window and a wide range of intervention types (e.g., enrichment, acceleration, differentiation, ability grouping, cognitive and socioemotional programs), outcome domains, and study designs. Data extraction with the validated codebook is currently underway. The complete list of included studies is openly accessible through the review's Covidence project and will be shared via the OSF repository together with the full dataset.

Expected results and contribution

By the time of the conference, screening and data extraction will be complete, and we will present the full results of the review. Specifically, the presentation will offer: (a) a comprehensive evidence map of school–based intervention research with gifted K–12 students published between 2010 and 2026, showing which interventions, outcomes, and populations have been studied and where important gaps remain; (b) a detailed methodological profile of the field, quantifying the relative prevalence of RCTs, quasi–experimental, single–group, and single–case designs, and appraising the

internal and external validity of group–design studies against WWC standards; and (c) an analysis of effect size reporting practices as an indicator of the transparency and quantitative maturity of the field, including the proportion of studies providing sufficient data for future meta–analytic synthesis. These findings will provide the first systematic and comparative assessment of the causal evidence base in gifted education over the last fifteen years. The results are expected to be of direct interest to researchers, practitioners, and policymakers: they will indicate whether and to what extent the evidence produced in the recent literature is sufficiently robust to support reliable inferences about intervention effectiveness, identify priority areas for future primary research, and offer concrete methodological recommendations for strengthening the design and reporting of intervention studies with gifted learners.

Keywords: Gifted education; Educational interventions; Scoping review; Experimental designs; Quasi–experimental designs; Evidence–based education

Gendered Recognition of Giftedness in Italian Primary Schools: Teachers' Attributions, Hidden Potential and Educational Equity

Erika Daria Torello – Eufrasio Pérez Navío – Enrico Bocciolesi

Introduction

The recognition of giftedness and high cognitive potential represents a central issue for contemporary education systems, not only in terms of teaching effectiveness, but also in relation to equity. Talent, indeed, cannot be regarded solely as a stable individual characteristic independent of context: its development also depends on educational opportunities, adult expectations, school cultures, and the tools through which potential is observed, interpreted, and supported. From this perspective, the transition from latent potential to the full expression of talent is not automatic, but requires timely identification processes, learning environments that are sensitive to individual differences, and adequately trained teachers.

International literature on gifted education has long highlighted the complexity of the processes involved in identifying gifted students, and several multidimensional models have contributed to moving beyond an exclusively psychometric view of giftedness. These models emphasise the interaction among cognitive abilities, creativity, motivation, family context, school environment, and educational opportunities. At the same time, numerous studies have shown how talent may remain invisible when it does not correspond to dominant representations of school excellence. This risk appears particularly relevant in the case of girls and young women, whose giftedness may be underestimated or recognised later due to gender stereotypes, social expectations, and modes of expression that are less immediately associated with the traditional image of the genius, brilliant, or cognitively precocious student.

This contribution is situated within this framework and examines the case of Italian primary school, a context in which gifted education still lacks a structured national system for early identification, as well as compulsory teacher training pathways, and where significant heterogeneity in school practices persists.

From this perspective, the issue of gender takes on particular relevance. The problem does not concern the actual distribution of giftedness between girls and boys, which cannot be inferred from teachers' perceptions alone, but rather the way in which potential is recognised, named, and reported. In other words, the central question is not "who is gifted?", but "who is more easily seen as gifted?". This distinction makes it possible to analyse giftedness not only as an individual phenomenon, but also as a relational and institutional one, influenced by teachers' implicit representations and by the often unspoken criteria through which schools interpret talent.

Data Used

This contribution is based on a sample of 225 Italian primary school teachers who took part in a questionnaire on knowledge, perceptions, experiences, and training needs related to gifted education. A specific area of analysis concerned gender attributions, with the aim of understanding whether giftedness was associated more frequently with boys, girls, or both.

In addition to the quantitative data, three focus groups with Italian primary school teachers were considered, as they were useful for further exploring the representations that emerged from the questionnaire. The contribution also refers to data from LabTalento at the University of Pavia.

Methodology

The study adopts a mixed–methods design. For the quantitative component, the frequencies related to gender attributions of giftedness were analysed, distinguishing between inclusive responses, referring to both boys and girls, and exclusive responses, referring only to boys or only to girls. The ratio between "boys only" and "girls only" attributions was interpreted not as an indicator of the actual distribution of giftedness, but as a synthetic measure of the representations declared by teachers.

For the qualitative component, the focus groups were analysed through thematic analysis, aimed at identifying recurring patterns and core meanings in teachers' discourse. The LabTalento data and the author's contextual observations were used exclusively as elements of empirical contextualisation, not as primary data nor as a basis for statistical inference.

The research involved only adults, on a voluntary basis and with informed consent. No sensitive data concerning minors were collected, and no diagnoses of giftedness were made.

Results and Discussion

The quantitative results show that most teachers recognise giftedness as a characteristic that may potentially be present in both boys and girls. The "mixed" category is, in fact, the most frequent response and suggests a formally inclusive representation of giftedness. However, the analysis of exclusive attributions reveals a critical element: when giftedness is associated with only one gender, it is attributed more frequently to boys. The ratio between exclusively male and exclusively female attributions is approximately 4:1, indicating a clear asymmetry in the representations declared by teachers.

This finding does not measure the actual distribution of giftedness between males and females, but signals a possible bias in recognition processes. In other words, some teachers appear to associate the gifted profile more easily with males. This result is consistent with international literature on gender bias in gifted education, according to which girls and young women with high potential may be less visible when they adopt behaviours that are more aligned with school expectations, show greater social adaptation, or do not express giftedness through immediately recognisable forms of achievement.

The integration of LabTalento data strengthens this interpretation. In the sample of children referred to the centre in Pavia, the subjects identified as having high potential are predominantly male. This finding, too, does not indicate a higher male prevalence of giftedness, but suggests an imbalance in the processes of referral and access to specialist assessment. It is possible that boys are more easily identified when they display distress, boredom, divergent behaviours, or difficulties in adaptation, whereas girls remain less visible when they express their potential through accuracy, commitment, responsibility, mature language, or positive academic performance. Consistently, also in the school contexts directly known to the author and for which reliable information is available, the pupils formally assessed as gifted are predominantly male. This element has exploratory and non–generalizable value, but reinforces the need to critically examine recognition processes.

The qualitative data from the focus groups allow this difficulty to be explored in greater depth. Teachers often describe gifted students through traits such as speed, curiosity, logical ability, advanced learning, rich language, and ease in completing tasks. However, ambivalences also emerge: some teachers associate giftedness with behavioural difficulties, boredom, disengagement, isolation, or relational difficulties; others struggle to distinguish between giftedness, high achievement, neurodevelopmental disorders, or special educational needs. This definitional uncertainty may make some profiles more visible and others less recognisable.

Within this framework, gender acts as a further interpretative filter. Gifted girls may be less recognisable not because they are less present, but because they are more frequently associated with behaviours of adaptation, commitment, and conformity to school rules. If these characteristics are not accompanied by explicit distress or divergent behaviours, they risk not triggering a request for further investigation. By contrast, boys who display cognitive precocity, intellectual liveliness, or behaviours misaligned with school expectations may be more easily noticed and associated with the image of the brilliant or exceptional student.

The issue therefore concerns the way in which schools define and recognise excellence. If talent is identified only through observable results, high performance, or distinctive behaviours, the education system risks identifying mainly those students whose potential is already visible. An equity–oriented perspective, instead, requires questioning the processes that precede achievement:

who is observed, who is referred, which behaviours are interpreted as indicators of high potential, and which students remain outside recognition pathways.
Teachers' training needs therefore concern not only theoretical knowledge of giftedness, but also the ability to observe different profiles, including those that are less aligned with the traditional image of the high–achieving child. In this sense, pedagogical observation tools can help make subtle signs visible, such as ethical sensitivity, emotional depth, atypical interests, creativity, perfectionism, dialogue with adults, divergent thinking, and performance anxiety, without turning the teacher into a clinical evaluator. Gender bias thus emerges as a central issue of educational justice, because it affects the possibility for every student to be seen, understood, and supported in their potential.

Conclusions

The study contributes to the debate on giftedness, early identification, and educational equity by proposing an interpretation of talent as a relational and contextual phenomenon. The results suggest that, in Italian primary school, giftedness is generally recognised by teachers as potentially present in both genders; however, when exclusive attributions emerge, the gifted profile is associated more frequently with boys than with girls. This asymmetry indicates the presence of possible gender bias in the processes of recognising potential.
The integration with LabTalento/Pavia data reinforces the relevance of the problem: also in pathways of access to specialist assessment, many more boys than girls appear to be identified. This finding does not demonstrate a higher male prevalence of giftedness, but suggests that referral and recognition processes may be influenced by the greater visibility of male distress and the lower visibility of female adaptation strategies.
The main educational implication concerns the need to strengthen teacher training in gifted education, with particular attention to gender stereotypes, non–conventional manifestations of high potential, and profiles at risk of remaining invisible. Effective training should help teachers distinguish between academic achievement, adaptation, behaviour, and potential, preventing giftedness from being recognised only when it takes culturally expected forms.
In conclusion, promoting equity in gifted education means critically examining recognition processes. Talent does not emerge merely because it exists, but because it is observed, understood, and supported. In the absence of a trained and aware pedagogical gaze, potential may remain hidden, especially in girls and in students who do not correspond to dominant representations of giftedness.

Keywords: Giftedness; Gender bias; Primary school; Educational equity

SESSION 8. EVALUATING UNIVERSITY PATHWAYS IN ITALY: INTEGRATED DATA, EDUCATIONAL CHOICES, STUDENT SUPPORT AND LABOUR MARKET TRANSITIONS

ORGANIZERS: UNIVERSITY OF PIEMONTE ORIENTALE – UNIVERSITY OF BOLOGNA
COORDINATORS: CARMEN AINA – GIUSEPPE PIGNATARO
6TH NOVEMBER: 8.30 A.M. – 10.30 A.M. {AULA MAGNA – RESEARCH 16}

University Scholarships and Academic Success in Italy: Evidence from National Administrative Data

Carmen Aina – Raffaello Bronzini – Augusto Cerqua – Ivan Etzo

Introduction

Financial constraints can affect higher education decisions at several stages. They may discourage university enrolment, reduce the time students can devote to study, delay credit accumulation, increase dropout risk, and limit progression to postgraduate education. From a human capital perspective, need–based financial aid can therefore improve both equity and efficiency: it broadens educational opportunities for students with limited resources while reducing the loss of talent and public investment associated with delayed or incomplete university careers.

A large international literature shows that grants can improve college access, persistence, and degree completion, especially when aid is sufficiently generous, predictable, and targeted at financial need (Castleman and Long, 2016; Denning et al., 2019; Herbaut and Geven, 2020). Evidence also indicates that programme design matters. Eligibility rules, merit requirements, payment timing, information barriers, and administrative complexity can alter take–up and student responses (Page and Scott–Clayton, 2016; Minaya et al., 2022). Italian studies generally identify positive effects on persistence, credit accumulation, and graduation, but much of the available evidence concerns individual universities, specific regions, or students close to an eligibility threshold (Mealli and Rampichini, 2012; Modena et al., 2020; Facchini et al., 2021; Rattini, 2023). National evidence covering several dimensions of academic progression remains comparatively limited.

This study examines the relationship between university scholarship receipt and academic success in Italy using integrated national administrative records. The analysis focuses on four outcomes that capture different stages of a university career: annual university credits, completion of a three–year bachelor programme within the legal duration or within one additional year, and immediate enrolment in a master's programme. We also assess whether the estimated relationships differ by gender, STEM enrolment, and geographical mobility for study.

The Italian setting is particularly relevant. University right–to–study policies combine a national regulatory framework with regional implementation. Scholarships are allocated mainly on the basis of economic need, subject to academic progress requirements, and interact with tuition exemptions, housing support, and other services. Recent increases in national and European funding have expanded scholarship coverage and amounts. Evaluating whether scholarship recipients progress more regularly is therefore important not only for distributional policy, but also for the allocation and operational design of higher education resources.

The study contributes in two ways. First, it uses population–level administrative data rather than information from a single institution or territory. Second, it considers progression, completion, continuation, and student heterogeneity within a common empirical framework. Because scholarship allocation is not random, the analysis does not claim to identify a definitive causal effect.

Instead, it estimates adjusted differences between recipients and observably similar non–recipients and explicitly discusses the remaining selection problem.

Data used

The analysis uses integrated administrative records covering Italian university careers from academic years 2018/19 to 2023/24. These data allow us to reconstruct enrolment histories, credit accumulation, degree completion, progression to master's programmes, scholarship receipt, tuition exemption status, secondary school background, and several individual and programme characteristics.

The main analytical sample includes students enrolled in academic year 2022/23 who obtained an upper secondary school diploma between 2020 and 2022 and were enrolled in a three–year bachelor programme. We restrict the comparison group to students receiving a full or partial tuition exemption. This restriction does not make recipients and non–recipients identical, but it excludes many students with clearly different economic profiles and creates a more relevant comparison group. The resulting sample contains 321,223 students: 98,468 scholarship recipients and 222,755 non–recipients. Students entered university in 2020, 2021, or 2022 and were therefore observed at different stages of the bachelor cycle in 2022/23.

The credit outcome measures the number of university credits earned during the relevant academic year. Completion outcomes are evaluated for the 2020–entry cohort, which contains 74,245 students, including 25,976 recipients. We measure whether students completed the bachelor degree within three years and within four years. The continuation outcome indicates immediate enrolment in a master's programme after bachelor completion. Geographically mobile students are defined as students enrolled in a university located in a region different from their region of residence. STEM programmes include scientific, technological, engineering, mathematical, agricultural, and related quantitative fields.

Methodology

We first compare unadjusted mean outcomes for scholarship recipients and non–recipients. We then estimate multivariate linear regression models for continuous and binary outcomes; coefficients for binary outcomes are reported as percentage–point differences. All specifications use heteroskedasticity–robust standard errors.

The covariate set includes age, gender, upper secondary school track, final diploma grade, university type, field of study, geographical area, and available information on fee exemptions and previous academic conditions. The models therefore compare students with similar observed demographic, educational, and university characteristics. We estimate the models for the full sample and separately for women, men, STEM students, and geographically mobile students.

Identification relies on a selection–on–observables assumption: after conditioning on included covariates, remaining scholarship status differences are assumed not to be systematically related to potential academic outcomes. This assumption is demanding because scholarship eligibility combines economic and merit criteria, while administrative records may not fully capture motivation, family support, employment during study, or other factors affecting both benefit receipt and performance. We consequently interpret the estimates as conditional associations that are consistent with an effect of financial support, rather than as conclusive causal estimates.

Results and discussion

Descriptive evidence shows a consistent advantage for scholarship recipients. Across the full sample, recipients earn 49.9 credits on average, compared with 43.9 among non–recipients. The raw difference is positive at each study stage and is largest in the third year. Among 2020 entrants, 67.2% of recipients complete the bachelor degree within three years, compared with 50.2% of non–recipients. Immediate master's enrolment is 36.3% among recipients and 28.4% among non–recipients. These gaps are substantial, but they may partly reflect differences in prior achievement, eligibility conditions, and other observed or unobserved characteristics.

Regression adjustment reduces some raw gaps but leaves a large and precisely estimated scholarship coefficient for every outcome. Scholarship recipients earn 4.72 additional credits in the academic year, conditional on the observed covariates. The estimated difference increases over the university cycle, from 4.21 credits in the first year to 5.03 in the second and 6.35 in the third. The third–year estimate corresponds to approximately 12.3% more credits relative to the comparison–group mean. This increasing pattern is consistent with the idea that financial support helps sustain academic continuity as students progress, although the model cannot isolate the precise mechanism.

Table 1. Adjusted scholarship coefficients for selected outcomes

Outcome	All	Women	Men	STEM	Mobile
Annual credits	4.72	4.58	4.95	5.69	3.41
Third–year credits	6.35	5.93	7.12	7.61	6.20
Degree within 3 years (p.p.)	24.40	23.80	25.42	23.41	22.03
Immediate master's enrolment (p.p.)	14.01	12.73	16.29	16.98	12.50

Notes: Coefficients from separate multivariate linear regressions with heteroskedasticity–robust standard errors. All reported coefficients are statistically significant at the 1% level. Binary–outcome coefficients are expressed in percentage points.

The completion estimates are especially large. Conditional on observed characteristics, scholarship receipt is associated with a 24.40 percentage–point higher probability of completing the bachelor degree within three years and a 22.28 percentage–point higher probability of completing within four years. Recipients are also 14.01 percentage points more likely to enrol immediately in a master's programme. The results therefore extend beyond annual performance: scholarship receipt is associated with a more regular trajectory, earlier degree completion, and greater continuation to the second cycle.

The subgroup estimates show that the positive relationship is not concentrated in one category of students. The annual credit coefficient equals 4.58 for women and 4.95 for men. It is larger among STEM students, at 5.69 credits, and smaller but still substantial among geographically mobile students, at 3.41 credits. At the third year, the corresponding estimates rise to 5.93 credits for women, 7.12 for men, 7.61 for STEM students, and 6.20 for mobile students. Completion and master's enrolment coefficients also remain positive and statistically significant in every subgroup.

Several interpretations are consistent with these findings. Financial support may reduce paid work during term, alleviate liquidity constraints, increase access to study materials and housing, and allow students to organise their studies more continuously. Merit requirements may also create incentives to maintain regular progress. At the same time, selection remains a plausible part of the explanation: students who satisfy scholarship conditions may differ in motivation, information, planning ability, or family circumstances that are not fully observed. The larger STEM coefficients may reflect the higher time intensity and lower work–study compatibility of these programmes, but the present data do not directly test this mechanism. The smaller average credit coefficient among mobile students may indicate that scholarships only partially offset housing, relocation, and adaptation costs.

From a policy perspective, the findings suggest that scholarships may generate benefits that go beyond redistribution. Faster progression and higher completion can reduce educational waste and

improve the return on public and private investment in higher education. However, the effectiveness of financial aid also depends on timely payment, clear eligibility rules, simple application procedures, and coordination with tuition exemptions, accommodation, and student services. Expanding nominal scholarship coverage without addressing delays or complementary costs may weaken the capacity of the policy to support regular study.

Conclusions

Using national administrative data, this study documents a strong and systematic relationship between university scholarship receipt and academic success in Italy. Recipients earn more credits, are more likely to complete a three–year degree on time or within one additional year, and are more likely to continue immediately to a master's programme. Positive coefficients emerge for women and men, STEM students, and geographically mobile students, although their magnitude differs across groups and outcomes.

The evidence supports the view that student financial aid can contribute simultaneously to educational equity and system efficiency. Policies that reduce financial constraints can help capable students sustain progress and complete their studies, limiting delays, dropout risk, and loss of human capital. The results also point to the importance of stable funding, timely disbursement, accessible information, and integration between scholarships, tuition exemptions, accommodation, and other support services.

The main limitation is the non–random allocation of scholarships. Rich covariates and a restricted comparison group improve comparability, but they cannot rule out selection on unobserved characteristics. Future work should exploit eligibility thresholds, policy changes, or longitudinal designs that provide stronger counterfactual identification. Further analyses should also examine scholarship effects on the initial university enrolment decision, interregional mobility, dropout, student employment, and longer–term labour–market outcomes.

Acknowledgements

This study was developed within the Italian Ministry of University and Research 2025–2027 Spending Analysis and Evaluation Plan. The authors thank Raffaello Bronzini, Luisa Lanza, Patrizio Memé, Claudia Pizzella, Francesco Salomone Marino, Paolo Turchetti, and Luca Cotti Zelati for their collaboration. The usual disclaimer applies.

Keywords: Student financial aid; Academic progression; Degree completion; administrative data

What if Gulliver Does Not Travel? Social Comparison, Ordinal Rank, and Transition into Higher Education

Andrea Lopes

Introduction

Humans consistently engage in social comparisons among themselves when interacting in groups. In psychology, a large body of research has shown that individuals form beliefs about their own ability, achievements, and social standing through comparisons with others, especially with peers who are perceived as similar or as psychologically close (Brown et al., 1992; Festinger, 1954; Gibbons et al., 1994; Taylor & Lobel, 1989; Tesser, 1988). In educational settings, such comparison processes are particularly important due to students' exposure to different information and signal, such as their peers' performance in exams through grades, classroom interactions, teachers' feedback, and the broader institutional environment. Consequently, students may form beliefs about their academic achievement and potential by positioning themselves relative to their peers along several dimensions, shaping their motivations, aspirations, effort, and subsequent educational choices. Within the economics of education, such phenomena are related to the "Big–Fish–in–a–Small–Pond"

hypothesis, according to which students with equal prior ability may form different academic self–concepts based on the interactions with their peers and on the ability composition of their reference group. In this sense, the ordinal position of a student in their reference group's ability distribution may carry signals that may improve students' performance (Marsh et al., 2008; Marsh & Parker, 1984; Murphy & Weinhardt, 2020; Yu, 2020). A growing body of literature documents the positive effects of ordinal rank on achievement and educational attainment in primary and secondary schooling (Carneiro et al., 2025; Claes et al., 2026; Denning et al., 2023; Dhuey, Elizabeth et al., 2025; Elsner & Isphording, 2017; Megalokonomou & Zhang, 2024), higher education (Delaney & Devereux, 2021; Elsner et al., 2021; Payne & Smith, 2020), labor market outcomes (Alonso, 2025; Chesney & Carrell, 2024), reduction in risky behavior (Chen et al., 2025; Comi et al., 2021; Elsner & Isphording, 2018), and improvements in mental health and non–cognitive outcomes (Kiessling & Norris, 2022; Kim & Liu, 2023; Pagani et al., 2021). Despite this recent and growing body of literature, less is known about whether the information students gather with their ordinal rank formed before higher education retains its signal along different transition margins into higher education. Using highly granular student–level microdata from the *Anagrafe Nazionale Studenti* (ANS) and the high school diploma grade as a proxy for cardinal ability from which the ordinal rank signal is generated, I contribute to this literature by leveraging quasi–random variation in the origin–group ability composition to estimate conditionally causal, within higher education sector ordinal rank effects on intensive and extensive margins of academic progression, over short– and long–run outcomes. The analysis is guided by the following hypotheses: (*i*) The ordinal rank signal (computed using the high school diploma grade) strengthens students' academic self–concepts and aspirations, stimulating further effort, and therefore improving students' performance at university; (*ii*) Given the formation prior to higher education transition, the ordinal rank signal should be on average equally predictive of students performance within each higher education sector, conditional on students' ability and characteristics; (*iii*) Building on the literature on rank saliency (Goulas & Megalokonomou, 2021; Megalokonomou & Zhang, 2024) and the different lags in transition timing within the two higher education sector, I hypothesize a *rank decay* mechanism, showing that the ordinal rank signal attenuates with enrollment delay; (*iv*) The high school rank signal may be heterogeneous in the information content in highly tracked systems. I hypothesize that such signal is stronger in academically oriented tracks (*lyceum*) rather than in technical or vocational tracks, reflecting differentiated signals for academic readiness and technical or vocational occupation–oriented skills, reflecting the differentiated functions historically embedded in the Italian high school track system since the 1923 Gentile Reform; (*v*) Additionally, I contribute the mixed evidence on gendered responses to ordinal rank (Delaney & Devereux, 2022, 2026; Elsner & Isphording, 2017) by analyzing, first, how effects differ when male and female students compare themselves within the overall origin–group ability distribution, and, second, by constructing within–gender rank measure, I measure how the effects differ when males and females compare themselves within gender–specific ability groups; (*vi*) Finally, in line with previous literature, I test the potential nonlinearities of ordinal rank effects by studying how they may differ along the percentile rank distribution.

Data

I leverage granular student–level microdata from the Italian Ministry of Education and Merit (MER) (*Ministero dell'Istruzione e del Merito*, *MIM*), called "*Anagrafe Nazionale degli Studenti e Laureati (ANS)*", covering the universe of students transitioning from high school to the Italian higher education system between 2010 and 2019. The final dataset comprises approximately 2.68 million student–level observations from 14,919 high schools and 92 universities, of which 81 are traditional universities and 11 are online universities. At the student–level, I observe students' demographic characteristics, including gender, country of birth, citizenship, municipality of residence, age at diploma, and age when enrolling at university. I also observe their detailed high school background and performance, including their final diploma grade, the year in which they obtained their diploma, unique high school identifiers, high school tracks and sub–tracks, and detailed geographical

characteristics. Such information allows me to reconstruct as granularly as possible the high school origin group in which I measure students' ordinal rank. Data also contains detailed information on students' higher education destinations, including university enrollment year, whether the student enrolls in a traditional or online university, the degree program, the degree class, the scientific field, the teaching municipalities, the access modalities, and the size of the enrolling cohort. Finally, my dataset allows me to merge institutions' and students' characteristics with their higher educational outcomes after enrollment. I can track students' credits, whether they reach relevant progression thresholds, their first–year dropout status, if they graduate, and their final degree mark, conditional on graduation. These students' outcomes permit me to study the impact of ordinal rank on both short– and long–run students' academic progression. My data allow me to reconstruct as granular as possible my percentile rank measure within high school origin peer groups, where I define the origin–group g as the set of students sharing the same high school s, the same (sub–)track k, the same graduation cohort c, and the same high school municipality M.

Methodology

The identification strategy exploits variation in students' relative position within high school origin peer groups, conditional on students' absolute ability, origin–group, and enrollment years fixed effects. Under the assumption that the residual variation is as good as random and that rank does not predict higher education sector assignment, the estimated coefficient on rank can be interpreted as conditionally causal within each higher education sector. The key idea is that, among students with similar final diploma grades and within the same origin group, differences in percentile rank arise from quasi–random variation in origin peer group ability composition. This allows me to isolate the role of rank as a psychological and educational signal of relative ability within each higher education sector. Specifically, within each origin–group g, I first rank students based on their proxy for ability (i.e., the diploma grade), such that the lowest–scoring student receives an absolute rank equal to 1 and the highest–scoring student receives an absolute rank of N_g, where N_g is the origin–group size. I then construct the percentile rank according to:

$$R_{i,g} = \frac{n_{i,g} - 1}{N_g - 1} \quad \text{s.t.} \quad R_{i,g} \in [0,1],$$

where 0 is assigned to the worst student and 1 to the best student; throughout the paper, I correct ties for the lower rank, following the literature, and I consider origin peer groups of size $N_g \geq 3$ to avoid systematically capturing extreme values too frequently. For each higher education sector $k \in K = \{T, O\}$, where T indicates *Traditional* and O the *Online* higher education sector, respectively, I estimate the following specification for students transitioning to traditional and online universities, separately:

$$y_{i,t(k)} = \alpha_k + \beta_k R_{i,g} + \gamma_k \mathbf{X}_i' + f_k(a_i) + \phi_g + \rho_t + \varepsilon_{i,t(k)}, \quad k \in K = \{T, O\},$$

where $y_{i,t(k)}$ denotes student's *i* academic outcome transitioning into higher education sector *k* in the enrollment year *t*. β_k is my coefficient of interest capturing the effect of a higher ordinal rank position $R_{i,g}$, conditional on a third–year polynomial for prior ability $f_k(a_i)$, on origin–group and enrollment year fixed effects, that is, ϕ_g and ρ_t, respectively. Finally, I two–way cluster the errors at the high school and university–class levels, which I define as the unique degree program × enrollment year × teaching municipality combination, allowing for arbitrary correlation in the error term among students sharing the same high school and among students entering the same university–class, where my outcomes of interest are formed.

Results and discussion

Relative to students transitioning to traditional universities, those who transition to online universities have a lower unconditional percentile rank (0.32 vs. 0.48), are on average older at enrollment (average age of 29.4 vs. 19.7), more likely to have repeated classes or to have non–linear high school trajectories, with an average age at diploma (20.2 vs 19.2), are predominantly males

(62% vs. 44%), have a lower average diploma grade (72.9 vs. 79.3), and they tend to transition more from technical (55.2% vs. 23.3%) and vocational tracks (12.1% vs. 6%), rather than an academically–oriented *lyceum* track (32.7% vs. 70.7%). I consider outcomes that differ along the following margins and time horizons: short–run intensive margins (*first–year credits*, *probability of reaching and surpassing the ANVUR credits threshold*); short–run extensive margin (*dropout probability at the end of the first year*); longer–run outcomes (*probability of graduation*, *final degree mark*). For students transitioning to traditional universities, I find that a 1 SD increase in my percentile rank measure, equivalent to moving up about 21 positions on average or 13 positions at the median, corresponding to an equivalent increase of roughly 11 points in the diploma grade, increases first–year credit accumulation by 0.51 credits, increases the probability of reaching the 40–credits ANVUR threshold by 0.7 percentage points (pp, hereafter), reduces the probability of dropping out by the end of the first by 0.36pp, increases the probability of graduating by about 0.65pp, and increases the final degree mark by around 0.21 points. The magnitude of the effects is modest, but my results are in line with the literature, highly stable, precisely estimated, and consistently signed. Conversely, for students transitioning to online universities, I find that the corresponding estimates are statistically indistinguishable from zero across all outcomes. To explain such results, I bring new evidence on a *rank decay hypothesis*: Focusing on students transitioning to traditional universities, I show that the ordinal rank signal weakens as the time between high school graduation and university enrollment increases, that is, the more students delay their transition into higher education. For most outcomes, the estimated rank effect becomes negligible after delaying enrollment for approximately one and a half to two years. I also find that the weakening pattern is stronger for males than for females, consistent with the composition of online university students. These results, robust across different specifications, suggest that new and relevant peer interactions external to the higher education sector weaken the salience of the original comparison group. This interpretation is consistent, on the one hand, with the absence of rank effects among students entering online universities, who are, on average, of lower ability, late learners, and predominantly male, and, on the other hand, with the literature on rank saliency. Moreover, when testing hypothesis (*iv*), I find that, among students transitioning to traditional universities, rank effects are strongest for *lyceum* students, whereas estimates for students from technical and vocational tracks are smaller and, in most cases, less precisely estimated, providing additional suggestive evidence on the null effects observed in online universities. Such patterns are also consistent with the idea that the diploma grade may not convey a uniform signal across a tracked school system. Specifically for my setting, in academically oriented tracks, the diploma grade may mainly signal readiness for further academic study. In technical and vocational tracks, instead, the same grade may capture a broader bundle of technical, practical, and vocational occupation–oriented skills, reflecting the differentiated functions historically assigned to secondary education since the Gentile Reform of 1923. Additionally, when testing hypothesis (*v*), I find that, for students entering traditional universities, female students appear to benefit more than male students from rank when measured in the overall distribution, generating potential gender gaps favoring female students. However, I also find that when students compare themselves within each gender–specific ability distribution, such differences are generally attenuated and, in some cases, the gender gaps favoring females close. Finally, I show the presence of nonlinear rank effects, showing that most of the effects on students' outcomes are concentrated in the tails of the ability distributions for students transitioning into traditional universities, in line with previous literature. Notably, for first–year credits, the 40–credit ANVUR threshold, graduation, and final degree marks, I find that students transitioning to traditional universities who are located in the lower ventiles perform worse than students around the median rank. Conversely, students located in the upper ventiles tend to perform better. For dropout, the signs are reversed, as expected. Consistent with previous literature, I find that effects display diminishing marginal returns in that the disadvantages associated with being in the lower parts of the rank distribution are larger and more persistent than the potential advantages for students whose ordinal position is above the median. That is, moving upward from

the bottom of the distribution is associated with relatively large gains in outcomes, but additional gains become smaller over much of the upper part of the distribution and are concentrated mainly at the very top. Generally, negative effects are more pronounced below the 7th–9th ventiles, while positive effects appear in the 19th and the 20th ventiles. An exception to this pattern appears when considering dropout probability, with statistically meaningful effects concentrated mainly at the two extremes of the rank distribution, that is, only in the lowest (1st) and highest (20th) quantiles.

Conclusions

In this paper, I study how students' ordinal rank in the high school ability distribution impacts students' academic progression when transitioning into higher education. Using administrative data on undergraduate students transitioning for the first time to traditional and online universities, and by comparing students with the same absolute ability who occupy different positions in their high school origin group ability distribution due to quasi–random variation in origin peer group composition, I estimate within higher education sector conditionally causal effects of high school ordinal rank on several higher education outcomes. My results show that, among students transitioning to traditional universities, a one standard deviation increase in percentile rank improves academic progression along both intensive and extensive margins. Higher–ranked students earn more first–year credits, are more likely to reach important progression thresholds, are less likely to drop out at the end of the first year, are more likely to graduate, and obtain higher final degree marks conditional on graduation, in line with previous literature. Although the magnitude of the effects is modest, ranging from approximately 0.01 to 0.08 SDs in absolute value, they are stable across specifications and consistently signed. Conversely, I find no statistically significant evidence that high school ordinal rank affects the same outcomes among students transitioning to online universities. I then analyze potential mechanisms and heterogeneities behind my main results. First, I find that the ordinal rank signal exhibits a decay pattern. Specifically, I show that ordinal rank effects are strongest for students transitioning immediately into (traditional) higher education after obtaining their high school diploma, and it weakens as the transition is delayed. This is consistent both with the idea that the original social comparison signal depreciates over time as former high school students move away from their high school reference group and acquire new information about their ability by interacting with new social groups external to secondary and tertiary education, and with the literature on rank saliency. Moreover, I find that the decay pattern is stronger for males than for females. Second, given that high school students coming from technical and vocational tracks constitute most of the students transitioning to online universities, I study differences in ordinal rank effects by high school track. I show that rank effects are strongest among students from lyceum tracks and weaker among students from technical and vocational tracks. Third, I study heterogeneous effects by gender using two different reference groups. On the one hand, I analyze whether the effects of gaining a higher ordinal position in the full origin–group ability distribution differ across genders. On the other hand, I construct a within–gender rank measure where I rank students within gender–specific origin–groups. My results suggest that the relevant comparison group differs across male and female students in traditional universities. Female students respond more strongly to their position in the overall high school ability distribution, generating gender gaps in estimated rank effects favoring females, whereas male students' outcomes become more responsive when rank is measured within same–gender peer groups. Fourth, I study non–linear rank effects, showing that rank effects in traditional universities are concentrated in the lower and upper tails of the rank distribution. Conversely, for students transitioning to online universities, my estimates remain close to zero across the full distribution, lending further support to my main results. Overall, my findings suggest that social comparison signals formed before university can shape later educational outcomes, but only when the original social comparison reference group remains psychologically, institutionally, and temporally relevant. In this sense, the results point to the importance of reducing enrollment delays to higher education and supporting

timely transitions to preserve the human capital premium associated with favorable higher relative position signals.

Keywords: Ordinal Rank; Higher Education; Social Comparison; Online Education

Unequal pathways? Migration background and university enrolment in Italy

Andrea Priulla – Fabio Aiello

Introduction

Educational systems worldwide have witnessed a significant expansion in higher education opportunities over the past decades. However, this expansion has not eliminated inequalities in access to university, which continue to vary systematically according to students' social origin, educational trajectories, and migration background. In particular, students with a migration background often face specific barriers in the transition from high school to university, making university enrolment a crucial stage at which broader educational inequalities may be reproduced.

Research has shown that these disparities are not solely attributable to migrant background itself, but rather result from the intersection of multiple factors, including socioeconomic status, parental education, family resources, educational aspirations, language proficiency, and structural features of educational systems, such as their degree of stratification and the availability of alternative educational pathways (Azzolini et al., 2012; Azzolini & Barone, 2013). Recent evidence confirms these patterns, suggesting that migration background interacts with social and institutional dimensions in shaping students' educational opportunities (Griga & Hadjar, 2014; Trappolini et al., 2025).

Italy represents a particularly relevant context, as the growing presence of students of migrant origin has increasingly raised questions about their educational integration and their access to post–secondary opportunities. Nevertheless, existing studies rely on limited survey data or data from single institutions, as administrative and longitudinal data have been lacking. For instance, using survey data, Mantovani et al. (2018) linked low enrolment rates among migrant students to early selection mechanisms. In Italy's stratified school system, students from migrant backgrounds are more likely to attend technical and vocational tracks, which offer fewer pathways to university than *licei* (Ballarino & Panichella, 2021). Paba and Bertozzi (2015) found that a *liceo* diploma increases university enrolment more for native than migrant students. More recently, Giudici et al. (2023) briefly described this gap using administrative data.

This study investigates inequalities in the school–to–university transition in Italy, focusing on high school graduates with and without a migration background. Italy represents a particularly relevant context, as the growing presence of students of migrant origin has increasingly raised questions about their educational integration and their access to post–secondary opportunities. Using a linkage of administrative data, we consider not only students' own origins but also their parents', allowing for a more detailed assessment of heterogeneity within the migrant background. Using logistic regression models, we examine differences in the likelihood of university enrolment, paying particular attention to how socioeconomic status and high school background shape educational trajectories differently across students' migration backgrounds.

Data

The analysis relies on a deterministic linkage of longitudinal microdata from INVALSI and the *Anagrafe Nazionale Studenti* (ANS). This linkage allows us to study the school–to–university transition in greater detail than has been possible before, given the limited availability of longitudinal administrative data. A key advantage of this dataset is that it enables examination, at the individual

level, of the role of socioeconomic status, which is not available in the ANS database, and of prior high school performance in shaping students' educational trajectories.
We focus on the graduate population from the 2021/22 academic year, corresponding to 497,019 graduates. For these students, we examine the likelihood of enrolling in an Italian university in the academic year following graduation, that is, in 2022/23.
To capture students' migration background, we construct a categorical variable, hereafter referred to as *student origin*, that combines information on students' place of birth with that on both parents' places of birth. The variable includes the following categories:

- *First–generation migrants*: students born abroad with both parents born abroad.
- *Second–generation migrants*: students born in Italy with both parents born abroad.
- *Native students*: students born in Italy with both parents born in Italy.
- *Mixed – Foreign Born*: students born abroad with one Italian–born parent and one foreign–born parent.
- *Mixed – Italian Born*: students born in Italy with one Italian–born parent and one foreign–born parent.

Students for whom information on the place of birth of the student or of at least one parent is missing are excluded from the analysis.
The final population under examination is described in Table 1, showing marked differences in enrolment rates. Overall, native students display the highest enrolment rate, followed by Italian–born from mixed parents and second–generation students, while first–generation and foreign–born students from mixed parents show the lowest levels. Across all groups, enrolment is highest among classical and scientific graduates and lowest among vocational ones, confirming the stratifying role of high school track.

Tab.1 – Descriptive statistics of students' population according to their origins, high school track, and university enrolment choice

	Hum/Sci		**Other Liceo**		**Technical**		**Vocational**		**Total**	
Student origin	**% Enr.**	**Tot**	**% Enr.**	**Tot**	**% Enr.**	**Tot**	**% Enr.**	**Tot**	**% Enr.**	**Tot**
1st generation	*78,4*	1573	*51,9*	2191	*35,9*	4180	*15,6*	3268	*39,1*	11212
2nd generation	*82,0*	4992	*58,5*	4920	*40,2*	8418	*18,4*	5393	*47,8*	23723
Mixed – For. Born	*72,9*	502	*47,4*	626	*36,4*	696	*13,3*	420	*43,3*	2244
Mixed – It. Born	*83,5*	9520	*59,1*	8302	*40,5*	9121	*15,6*	4524	*54,8*	31467
Native	*87,0*	121656	*63,4*	88033	*40,3*	107544	*15,6*	51773	*57,7*	369006
Total	***86,5***	**138243**	***62,5***	**104072**	***40,2***	**129959**	***15,8***	**65378**	***56,5***	**437652**

Methods
We estimate two logistic regression models to analyse the probability of enrolling at university after high school graduation. In addition to high school track, socioeconomic status (measured through the student's ESCS) and students' origins, the model includes: sex (female or male); macro–region of high school graduation (North, Centre, and South and Islands); high school regularity (1 if the student graduated without any delay or failure, 0 otherwise); student's score in the grade 13 INVALSI Italian test; share of students with foreign citizenship in the same high school classroom.
Model 1 includes an interaction between migration background and high school track.

Model 2 extends the analysis by including a three–way interaction between migration background, high school track, and socioeconomic status. This allows us to examine whether socioeconomic status shapes the school–to–university transition differently across students' migration backgrounds and tracks.

Results

Figure 1 shows the predicted probability of university enrolment from model 1, with all other covariates held at their baseline (mean) values. The results highlight clear differences by student origin, although these vary across high school tracks. Among lyceum graduates, native students show the highest probability of enrolment, while Mixed – Foreign Born students display the lowest. By contrast, among technical and vocational graduates, the pattern is partly reversed: first– and second–generation students show probabilities comparable to, or higher than, those of native students. These differences are embedded in a strong track–related gradient, with enrolment probabilities highest among lyceum graduates and much lower among technical and vocational graduates.

Fig. 1 – Model 1: probability of university enrolment by high school track and students' origins

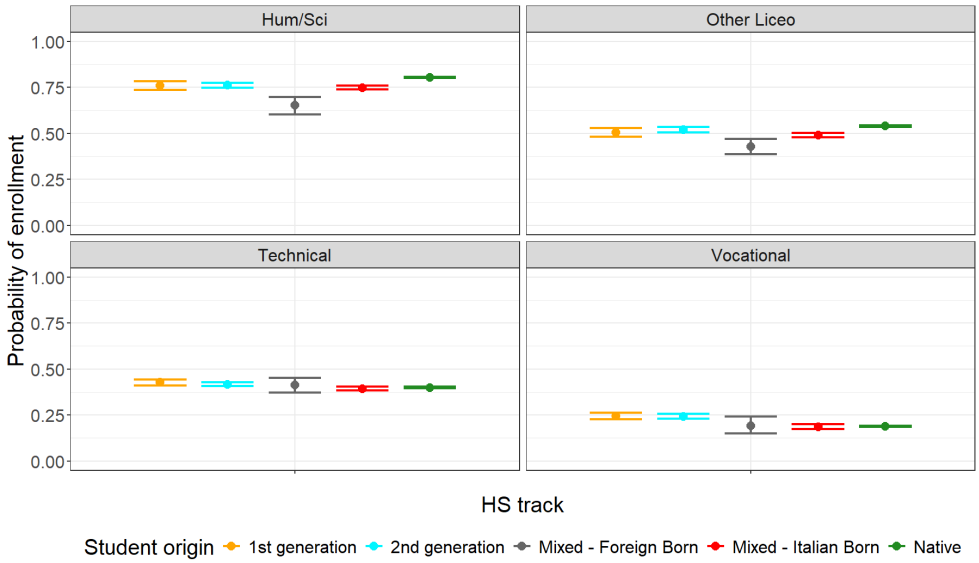


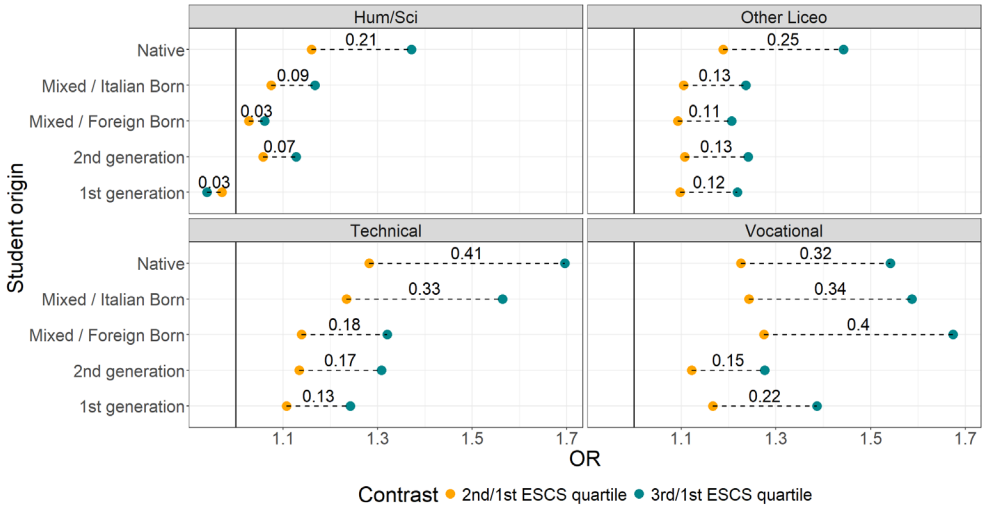
Figure 2 reports the odds ratios from Model 3 for university enrolment by students' socioeconomic status, comparing the second and third ESCS quartiles with the first quartile, separately by student origin and high school track. The results show that the role of socioeconomic status is generally stronger among native students, for whom the odds of enrolling at university increase markedly as ESCS rises. By contrast, the socioeconomic gradient appears weaker among first– and second–generation students, suggesting that differences in university enrolment within these groups are less strongly structured by family socioeconomic background. This pattern is particularly evident among lyceum and technical graduates, while results for vocational graduates are more heterogeneous.

Conclusions

This study provides new evidence on inequalities in the school–to–university transition in Italy by considering not only students' own migration background, but also parental origin. Using linked administrative data, we show that university enrolment varies substantially across origin groups and high school tracks. Native students display the highest overall enrolment rates, while first–generation and Mixed – Foreign Born students are generally less likely to enter university. However, these differences are not uniform across tracks. Among lyceum graduates, native students retain a clear advantage, whereas among technical and vocational graduates, first– and second–generation students show enrolment probabilities comparable to, or even higher than, those of native students.

The results also highlight that socioeconomic status shapes educational pathways differently across origin groups. The ESCS gradient is stronger among native students, whereas it appears weaker among first– and second–generation students, suggesting that the role of family socioeconomic background in university enrolment differs across groups. Overall, the findings point to the importance of analyzing migration background in greater detail, distinguishing between students' and parents' origins and considering how these dimensions interact with high school track and socioeconomic status. This approach reveals patterns of inequality that would remain hidden in broader comparisons between native and migrant–background students.

Fig. 2 – Model 2: odds ratios for university enrolment conditional on students' socioeconomic status — second and third quartiles — by high school track and student origin. The baseline category for socioeconomic status is the first quartile.



.

Keywords: Higher education; Online universities; Academic outcomes; Dropout rates

The Evolution of Comparative Advantage During Secondary School and University Field Choice: Evidence from Linked INVALSI and Administrative Data

Luca Bonacini – Carmen Aina – Chiara Mussida – Giuseppe Pignataro

Introduction

Women remain substantially underrepresented in Science, Technology, Engineering and Mathematics (STEM) fields despite their overall educational attainment exceeding that of men in most developed countries. Understanding the origins of gender differences in university field choice has therefore become a central question in the economics of education.

A growing literature investigates whether gender disparities in STEM enrolment reflect differences in academic skills, preferences, or educational trajectories developed before university entry. Recent studies show that a large share of the STEM gender gap can be traced to differences in secondary–school preparation and course choices (Card and Payne, 2021; Granato, 2023). Another strand of literature highlights the role of comparative advantage, arguing that students choose fields not only according to their absolute performance in mathematics but also according to their relative

strengths across subjects. Breda and Napp (2019) show that gender differences in comparative advantage between mathematics and reading can explain a substantial fraction of the STEM gender gap. However, more recent evidence suggests that the explanatory power of comparative advantage may be considerably smaller than initially estimated (Wan et al., 2023).

While existing studies typically measure comparative advantage at a single point in time, much less is known about how comparative advantage evolves during secondary school and whether this evolution shapes subsequent university choices. This question is particularly relevant because educational aspirations, academic identities, and subject–specific competencies develop throughout adolescence.

This paper provides new evidence on the relationship between comparative advantage and university field choice using linked administrative data from Italy. Combining INVALSI test scores collected in the second and fifth years of upper secondary school with administrative university enrolment records, we investigate whether mathematics achievement and comparative advantage predict enrolment in STEM programmes and whether their predictive power changes over the course of secondary education.

To our knowledge, this is the first study to examine how the evolution of comparative advantage during secondary school is associated with university field choice in Italy.

Data

The analysis relies on the integration of two large administrative sources.

The first source is the INVALSI student database, which contains standardized test scores in mathematics and Italian language collected at different stages of schooling. We use information from students tested during the second year of upper secondary school (Grade 10) and during the final year of upper secondary school (Grade 13).

The second source is the Italian National Student Registry (Anagrafe Nazionale Studenti), which provides information on university enrolment and degree programmes.

The linked dataset includes approximately 243,000 students who completed upper secondary education and subsequently enrolled in university programmes. We classify degree programmes into STEM and non–STEM fields. Following a more restrictive definition commonly used in the literature, we also distinguish between hard STEM programmes (engineering, mathematics, physics, statistics, computer science, and related fields) and soft STEM programmes.

The final sample consists of 56.4% female students and 43.6% male students. Approximately one–third of students enrol in STEM programmes, while around 22% enrol in hard STEM programmes.

Methodology

We construct standardized measures of mathematics achievement and comparative advantage at both observation points.

Mathematics achievement is measured using standardized mathematics test scores. Comparative advantage is defined as the difference between mathematics and Italian–language test scores and then standardized within cohort.

Our empirical strategy proceeds in three steps.

First, we document gender differences in mathematics achievement, language achievement, and comparative advantage during secondary school.

Second, we estimate linear probability models relating STEM enrolment to gender, mathematics achievement, and comparative advantage. We evaluate the extent to which these measures account for the observed gender gap in STEM enrolment.

Third, we investigate heterogeneity along the skill distribution by estimating gender differences in STEM enrolment across deciles of mathematics achievement and comparative advantage. This allows us to assess whether gender disparities are concentrated among low–performing or high–performing students. Finally, we estimate models including achievement and comparative advantage measured both in Grade 10 and Grade 13 to evaluate whether later assessments contain additional information beyond earlier measures.

Results and Discussion

Several findings emerge.

First, sizeable gender differences are already visible during secondary school. Female students score substantially lower in mathematics and slightly higher in Italian language. As a consequence, females exhibit a markedly lower comparative advantage in mathematics relative to language. The gender gap in comparative advantage reaches approximately 0.55 standard deviations in Grade 13 and exceeds 0.60 standard deviations in Grade 10.

Second, both mathematics achievement and comparative advantage strongly predict STEM enrolment. A one–standard–deviation increase in mathematics performance is associated with a sizeable increase in the probability of choosing a STEM degree. Comparative advantage also plays an important role, although its explanatory power is smaller than that of mathematics achievement.

Third, mathematics achievement explains a larger share of the STEM gender gap than comparative advantage. Controlling for Grade–13 mathematics scores reduces the gender gap in STEM enrolment by roughly 29%, whereas controlling for comparative advantage alone reduces the gap by approximately 24%. Similar patterns emerge when considering Grade–10 measures.

Fourth, important differences emerge between STEM subfields. For hard STEM programmes, mathematics achievement remains the dominant predictor, whereas comparative advantage plays a more limited role. By contrast, female students are slightly more likely than males to enrol in soft STEM programmes, suggesting that gender segregation is concentrated in the most mathematics–intensive fields.

Fifth, the relationship between skills and university choices varies substantially across the achievement distribution. Gender differences in STEM enrolment become larger among students with higher mathematics achievement and among students with stronger comparative advantage in mathematics. This finding suggests that the underrepresentation of women is particularly pronounced among students with the strongest academic profiles.

Finally, when measures from both Grade 10 and Grade 13 are included simultaneously, later assessments retain substantially greater predictive power. Nevertheless, both early and late measures remain statistically significant predictors of STEM enrolment. This result suggests that the development of academic skills and comparative advantage throughout secondary school continues to shape educational choices beyond initial differences observed at earlier stages.

Conclusions

This paper examines the role of mathematics achievement and comparative advantages in explaining gender differences in university field choice using linked INVALSI and administrative university records.

The results indicate that both absolute mathematics performance and comparative advantage are important predictors of STEM enrolment. However, mathematics achievement consistently explains a larger share of the gender gap than comparative advantage. Moreover, measures collected at the end of secondary school are substantially more predictive than those observed earlier, highlighting the importance of skill development during adolescence.

The findings contribute to the literature on gender disparities in STEM by showing that educational trajectories throughout secondary school play a crucial role in shaping university choices. They also suggest that policies aimed at reducing gender segregation in STEM should focus not only on improving girls' mathematics performance but also on understanding how comparative advantage evolves during the secondary–school years and influences educational decisions.

Keywords: Comparative Advantage; Gender Gap; Math; Stem

Beyond prediction: uncovering dropout patterns with machine learning in Italian schools

Melisa L. Diaz Lema – Mariavittoria Ludovico – Tommaso Agasisti

Introduction

Preventing school dropout remains one of the most pressing challenges for education systems worldwide. Early school leaving carries well–documented long–term consequences: lower earnings, higher unemployment rates, worse health outcomes, and reduced civic participation (Oreopoulos & Salvanes, 2011; Lleras–Muney, 2005; De Witte et al., 2013; OECD, 2023). Despite growing recognition of the problem, dropout is still treated as a single, homogeneous phenomenon rather than as the outcome of distinct and divergent pathways. Students abandon school for fundamentally different reasons such as academic failure, disengagement, socioeconomic pressure, or sudden life events requiring very different responses (Janosz et al., 2000; Bowers & Sprott, 2012; Dupéré et al., 2015, 2018). A single risk score conflates students who need different support: the heterogeneity in dropout pathways is lost in aggregation, and so is the policy signal.

Background and Motivation

Early Warning Systems (EWS) have emerged as a promising tool for identifying at–risk students before dropout occurs, increasingly relying on high–dimensional machine learning models that outperform traditional statistical approaches in terms of predictive accuracy (Sansone, 2019; Márquez–Vera et al., 2016; Alam & Bowden, 2024). Academic achievement, absenteeism, behavioural issues, socioeconomic background, and school–level characteristics have all been identified as key predictors of dropout, reflecting the complex interplay between individual performance, family context, and the educational environment. Yet EWS show limited or no causal effect on actual dropout rates, pointing to a fundamental gap between prediction and effective intervention (Sletten et al., 2023; Haimovich et al., 2021; Canbolat, 2025). We argue this gap is rooted in the one–size–fits–all architecture of most EWS: knowing who is at risk, without understanding why, provides insufficient guidance for targeted action. This paper moves beyond a single risk score to uncover the heterogeneous pathways that drive early school leaving among Italian upper secondary students, drawing on a rich administrative dataset and an interpretable machine learning framework. Three research questions guide the analysis: (RQ1) What are the main drivers of predicted dropout risk among Italian upper secondary students? (RQ2) Can distinct typologies of dropout students be identified based on their individual risk–driver profiles? (RQ3) Does predictive performance vary systematically across these dropout typologies?

Data and Methodology

The study draws on population–level administrative records from the Italian Ministry of Education, merged with INVALSI standardised test data, covering students enrolled in Grade 9 of upper secondary school during the academic year 2021–22, in schools that recorded at least one dropout case[26]. The dataset encompasses four information domains: student background characteristics, prior academic performance and engagement, peer environment characteristics at the classroom level and school organisational features. The empirical strategy unfolds in three steps. First, a gradient boosting model (XGBoost) is trained to predict dropout probability on stratified training, validation, and test splits. Because dropout is rare and failing to flag a future leaver is more costly than a false alarm, positive cases are weighted in proportion to their rarity, hyperparameters are

[26] A structural feature of the study window is that INVALSI was not administered in 2019-20 because of the COVID-19 pandemic, leaving students who sat Grade 8 in that year (disproportionately retained students) without test scores or questionnaire responses. Rather than discarding or imputing these observations, we treat the missingness as potentially informative, adding explicit indicators of whether each value is observed so that the model can use them.

tuned to maximise the area under the precision–recall curve (AUCPR), and the classification threshold is set to maximise the F2 score on the validation set. SHAP values are then computed to decompose each student's predicted risk into the individual contribution of each observable characteristic, enabling the transition from global feature importance to individualised risk–driver profiles. Second, dropout students are clustered in SHAP space rather than on raw observables, using a partitioning around medoids algorithm with the number of clusters selected by average silhouette width, supplemented by the substantive interpretability of the resulting profiles where the criterion is close to indifferent. Retained and never–retained students are clustered separately, yielding two profiles among the retained and three among the never–retained. Clustering in SHAP space groups students by why they are at risk rather than who they are, producing profiles that are more clearly separated and directly interpretable in policy terms (Lundberg, Erion & Lee, 2018; Cooper et al., 2021). Third, the resulting profiles are characterised through observable characteristics, and predictive performance is compared across typologies to assess whether the model identifies all dropout students equally well.

Results

The analysis yields five distinct dropout profiles, structured around a fundamental divide between retained and never–retained students, who occupy entirely separate regions of risk–driver space.

Among retained students, two profiles emerge. The first one, *Persistent Academic Failure,* is characterised by a long and cumulative history of academic underperformance, with grade retention as the dominant risk driver. These students concentrate in technical and vocational school tracks and in contexts of pronounced socioeconomic disadvantage; intervention for this group could emphasize intensive academic support and remedial programmes. *School Mobility* students, despite sharing a history of retention, show comparatively lower predicted risk and come from less disadvantaged backgrounds, suggesting that school change may in some cases reflect an active re–engagement strategy rather than a further sign of deterioration. Among never–retained students, three profiles are identified. *Overage Disengagers* are older than expected for their grade level, often from immigrant backgrounds, with high absenteeism and low behavioural grades, suggesting a protracted process of disengagement that may have begun long before the observation window; mentoring, counselling, and closer family engagement may be most effective for this group. *Young Behavioral Risk* students are characterised by significant behavioural difficulties and are concentrated especially in technical school tracks, pointing to conduct–based disengagement; mentoring and psychological counselling could prove effective. The third profile is the *Unexpected Dropout* group: students who leave school without leaving any detectable trace in administrative records, presenting near–average grades, attendance rates, and test scores. The model assigns them very low predicted dropout probabilities, meaning that a substantial share of never–retained dropouts would be systematically missed by current Early Warning Systems. Understanding why this group leaves school and how to intervene would require data that go beyond what administrative records can capture.

Discussion and Conclusion

These findings carry important implications for both research and policy. The heterogeneity of dropout pathways demands heterogeneous responses: intensive academic remediation for students facing persistent academic failure, mentoring and family engagement for overage disengagers, and community–based or school–level strategies for those whose risk is contextually driven. The existence of an undetectable dropout group raises questions about the limits of administrative data, calling for complementary data collection strategies, including non–cognitive, qualitative, and relational information, that may capture signals invisible to current EWS architectures. More broadly, the results suggest that Early Warning Systems should move from a single–score logic toward a profile–sensitive framework, in which the nature of the risk, and not only its magnitude, guides the

institutional response. Ultimately, predictive accuracy alone is insufficient: the true value of an EWS lies in its ability to transform predictions into differentiated and meaningful action.

Keywords: School dropout; Dropout typologies; Machine learning; Supervised clustering

SESSION 3. INTERNATIONAL LARGE–SCALE ASSESSMENTS (ILSAS)
METHODS AND RESULTS
ORGANIZER: UNIVERSITY OF BATH
COORDINATORS: MARIA MAGDALENA ISAC – ANDRES SANDOVAL HERNANDEZ
6TH NOVEMBER: 8.30 A.M. – 10.30 A.M. {AULA A1.1 – RESEARCH 17}

What is civic knowledge? An in–depth analysis of the ICCS Civic Knowledge Framework and its evolution over time

Jan-Philipp Wagner – Francisco Gatica

The International Civic and Citizenship Education Study (ICCS) is an international comparative research project commissioned by the International Association for the Evaluation of Educational Achievement (IEA) and currently directed by the International Study Center located at the University of Bath (International Association for the Evaluation of Educational Achievement (IEA), 2025). ICCS aims at assessing civic and citizenship education (CCE) in different education systems around the world "to investigate the ways in which young people are prepared to undertake their roles as citizens" (Schulz et al., 2016: 1) by assessing eighth graders' civic knowledge as well as their political attitudes, their predispositions toward different aspects of society and civic principles, and their current and expected future civic engagement. To that end, the study comprises of a civic knowledge test, an international questionnaire and optional regional questionnaires administered to students as well as a questionnaire for teachers, a questionnaire for school principals, and a national contexts survey completed by national CCE experts (International Association for the Evaluation of Educational Achievement (IEA), 2025). ICCS is informed by the national curriculums of the participating education systems, national and international research interests and the educational policy priorities of national and international organizations. As such, the construct of civic knowledge (CK) that is positioned at the core of the study has evolved over time, from ICCS's precursor CIVED in 1999 to the latest iteration of the study in 2022.

Just like other international large–scale assessments (ILSAs) in education, such as PISA and TIMSS, ICCS obtains information on student learning outcomes at the level of the education system and facilitates comparisons over time and across education systems (Lietz et al., 2017). Lietz and colleagues (2017) show that ILSAs contribute to educational policy–making in three ways: (1) by comparing educational achievement across education systems and over time, ILSAs facilitate the measurement of educational quality; (2) by comparing educational achievement across subgroups (e.g. gender, socio–economic status, school type), ILSAs contribute to measuring educational equity; and (3) by monitoring and reporting achievement results, ILSAs help generate accountability in the education sector. As such, ILSAs can be considered as tools implemented to further, support, and/or reform educational policies based on the measurement of educational outcomes. In the case of ICCS, these educational outcomes are defined in the assessment frameworks, with the ICCS civic knowledge framework describing the understanding of civic knowledge as measured by the study.

The ICCS civic knowledge framework is developed by the ICCS International Study Center in collaboration with the study's stakeholders, namely IEA, the National Research Coordinators representing the national interests of the countries participating in ICCS, and international experts. Each of these stakeholders bring a combination of organisational, political and research interests to the development of the civic knowledge framework.

The civic knowledge framework defines the content domains and cognitive domains covered by the ICCS civic knowledge test. "Content domains describe areas related to civic and citizenship education

about which individuals may have developed knowledge and understanding" (Schulz et al., 2022, p: 27). The cognitive domains refer to the cognitive processes that students need to employ to demonstrate their civic knowledge as measured by ICCS.

Within the context of ICCS, the construct of civic knowledge lies at the centre of tension between continuity and innovation: on the one hand, the construct needs to remain to stable to facilitate comparisons over time; on the other hand, innovation is needed to ensure that the study continues to be relevant in a changing reality of CCE.

While maintaining consistency at its core, the ICCS civic knowledge framework has been changed and refined over time. While the precursor of ICCS, the IEA Civic Education Study (CIVED) from 1999, does not offer a concrete definition of civic knowledge (Torney–Purta et al., 2001), the ICCS 2022 civic knowledge framework defines civic knowledge as "demonstrable student achievement … that extends beyond their capacity to recall information and includes students' ability to reason with and apply their knowledge" (Schulz et al., 2022, p. 27). Moreover, the construct of civic knowledge has been expanded over the study cycles. For instance, the ICCS 2016 assessment framework added knowledge about environmental sustainability and awareness of the economy to the construct (Schulz et al., 2016). The ICCS 2022 framework then identified Education for Sustainable Development (ESD) and Global Citizenship Education (GCE) as distinct aspects of civic knowledge (Schulz et al., 2022).

An understanding of ICCS as an extension of educational policy allows us to consider the civic knowledge framework (and other documentation of the study) as a political artifact whose evolution can be studied using policy analysis.

Policy Analysis: Drivers of Change

This paper looks at ICCS through a policy analyst's lens: taking the ICCS civic knowledge framework as a policy artifact representing the interests of the stakeholders involved in ICCS, the paper applies a framework for policy analysis to assess the drivers of the evolution off the civic knowledge framework and the construct of civic knowledge within it. In the proposed framework, decisions to make changes to the civic knowledge framework (or not) can be motivated by political reasoning, theoretical developments, empirical advancements, and strategic factors.

To understand consistency of the construct over time, the concept of Path Dependence (PD) is applied to the ICCS civic knowledge framework. PD stipulates that past decisions make policies highly resistant to change (Pierson, 2000). In the context of ICCS, PD suggests that once "civic knowledge" was defined in a certain way, deviations from this definition became increasingly unlikely since the costs of change would outweigh its benefits. In ILSAs, conceptual continuity is further strengthened by stakeholders' interest in measuring change in educational outcomes over time, which requires stable measures and construct definitions.

To explore innovation and change of the construct, John Kingdon's multiple stream analysis (MSA) is applied to the civic knowledge framework. MSA suggests that three conditions must be met for a policy to change: (1) the presence of an identifiable problem; (2) the presence of a plausible solution to that problem; and (3) the presence of political will to solve the problem (Kingdon, 1995). Applying this logic to the ICCS civic knowledge framework suggests that a change of the definition of civic knowledge requires the existence of a problem (e.g., a need to systematically monitor ESD and GCE across education systems), the identification of a feasible solution (e.g., the inclusion of ESD and GCE in ICCS), and the support from key stakeholders, such as funders, IEA, national study centres, and researchers.

Methodology

To understand changes in the construct of civic knowledge over time, an in–depth content analysis of essential study documents is carried out, including the CIVED International Report (Torney–Purta et al., 2001) and the various ICCS civic knowledge frameworks (Schulz et al., 2008; Schulz et al., 2016; Schulz et al., 2022). The outcome of this document analysis is a detailed mapping of the components of the construct, highlighting continuities and changes.

To assess why certain changes to the civic knowledge framework have occurred, semi–structured interviews with key stakeholders in ICCS (e.g., the former lead of item development, research advisors at IEA, National Research Coordinators) will be conducted to identify the reasons that drove changes to the civic knowledge framework from study cycle to study cycle. Interviewees will be asked to reflect on the evolution of the framework, referring to specific changes and the likely reasons for them.

The interviewees' responses will be analysed in light of the aforementioned theories of policy change to generate insights into the decision–making process behind the evolution of the ICCS civic knowledge framework. The interview transcripts will be screened to identify the overarching theme (change vs. continuation), stakeholders involved in the decision–making process as well as likely reasons for the decisions.

Results and Discussion

As this is an ongoing study, concrete results are not yet available. However, it is expected that some initial results can be presented at the time of the conference (November 2026). Moreover, the study will ultimately contribute to the field of civic education by providing an in–depth mapping of the ICCS civic knowledge construct over time, accompanied by a comprehensive explanation of specific changes to the construct. Moreover, the paper will offer a theoretical and analytical framework for assessing the evolution of assessment frameworks that can be applied to other international large–scale assessments (ILSAs) in education, such as TIMSS and PISA.

Keywords: Civic Education; Civic Knowledge; Framework Analysis; Policy Analysis

Can Civic Education Compensate Social Inequalities in Civic Outcomes? Evidence from Chile and Comparative Large–Scale Assessments

Daniel Miranda – Jorge Matamala – Benjamín Zavala

Introduction and Problem Statement

The optimal functioning of a democracy rests on the premise that its citizens participate equitably in decision–making processes, show interest in public affairs, adhere to democratic principles, and develop attitudes consistent with life in a democratic society (Hoskins & Janmaat, 2019; Rouse et al., 2022; Schlozman et al., 2012). However, specialized literature consistently demonstrates that multiple background gaps—socioeconomic, gender, and ethnic—threaten this egalitarian ideal, directly affecting access to political voice, interest, civic knowledge, and adherence to fundamental values.

During childhood and adolescence, the family and the school constitute the two most relevant institutions in the process of political socialization, frequently operating under opposing logics. On one hand, the family acts as a channel for social and political reproduction: it directly transmits beliefs, attitudes, and civic skills, but also passes the advantages and privileges of its own structural position to subsequent generations. Inequalities in the social structure deeply mark family political socialization, translating into stratified ways of experiencing politics (Peña, 2015).

On the other hand, institutional education is considered the critical space for promoting democracy and training capable citizens, under the explicit assumption that it has the potential to compensate for background inequalities. Unlike family socialization, the school system generates a series of formal and informal activities that shape a common experience for all students attending the system. Within this framework, the compensation hypothesis emerges (Campbell, 2019), suggesting that a homogeneous school experience can level, moderate, or mitigate gaps caused by social origin. Under this assumption, two students from opposing socioeconomic positions should develop similar levels

of civic knowledge or participatory propensity when exposed to equivalent processes of school socialization.

However, classical sociology of education (Bourdieu, Coleman, Bernstein) has historically argued that this equalizing ideal is not met; rather, the school tends to operate as an agent that reproduces and validates background inequalities rather than correcting them. According to Hoskins & Janmaat (2019), the school system can fail in its leveling role through two mechanisms: *unequal provision* (generating differentiated and lower–quality teaching and learning opportunities for students from vulnerable environments) or *acceleration of differences* (when advantageous background conditions allow privileged students to better capitalize on and profit from universal learning opportunities).

Faced with this conceptual tension, this proposal addresses the problem through a dual empirical strategy: it evaluates the persistence of these gaps across different points in time within a highly segregated national context (Chile) and, simultaneously, contrasts these findings at a transnational level using data from the International Civic and Citizenship Education Study (ICCS). The objective is to determine whether school–based civic education compensates for, maintains, or accelerates background civic inequalities.

Theoretical Framework and Competitive Hypotheses

Contemporary research using Large–Scale Assessments (LSAs) is strictly divided into three empirical stances regarding the impact of school factors on background civic gaps:

The Mitigation (or Compensation) Hypothesis

This perspective argues that schooling reduces civic knowledge gaps by disproportionately benefiting disadvantaged students. However, previous findings are highly context–specific and depend on the school dimension analyzed. Regarding curricular design, Gainous and Martens (2012) demonstrated in the United States that a broad curriculum focused on a diversity of civic topics significantly increases political knowledge and internal efficacy only among students from low socioeconomic status (SES) families, serving as an effective equalizer. In terms of internal participation, Hoskins et al. (2021) identified a context–bounded compensatory effect in Nordic countries, specifically in Finland, where active civic participation within schools helped counteract socioeconomic disadvantages in the acquisition of citizenship competencies. Furthermore, regarding classroom social composition, Kudrnáč and Lyons (2018) found in the Czech Republic that low–SES students increased their civic knowledge to a greater extent when integrated into classrooms with a higher average SES than their more advantaged peers in the same environment.

The Maintenance Hypothesis

A substantial body of research contends that schools maintain rather than reduce civic inequalities. Although school factors exert a generalized positive impact on all students' performance, they fail to alter the underlying socioeconomic gap. Campbell (2008), using data from the CIVED (1999) study in the US, evidenced that an open and deliberative classroom climate distributes its benefits homogeneously across all strata, thus failing to reduce the gap between high– and low–SES students. This maintenance hypothesis finds strong backing in recent longitudinal and international studies. Persson (2015), combining Swedish panel data and transnational evidence, rejected the compensation hypothesis by showing that an open classroom climate increases factual political knowledge in the same magnitude for all students, perpetuating the initial distance. In the Czech context, Kudrnáč (2017) and Kudrnáč and Lyons (2018) confirmed that this factor is insufficient to offset the disadvantages associated with low SES. In the same vein, Hoskins et al. (2021) corroborated this with ICCS data (2009 and 2016), finding that an open climate reproduces pre–existing gaps in most Nordic countries, suggesting that universal school interventions raise overall performance without restructuring baseline inequality.

The Acceleration Hypothesis

Conversely, certain findings warn that some school characteristics can actively widen background disparities, predominantly benefiting those who already possess initial advantages. Hooghe and

Dassonneville (2011), using the Belgian Political Panel Study (BPPS), discovered that group projects favor students with higher prior political knowledge to a greater extent, amplifying the gap over time. Similarly, Gainous and Martens (2012) reported that *instructional breadth* (a wide variety of pedagogical methods) harms the political knowledge of low–income students while leaving privileged students unaffected, thus expanding the gap.

This acceleration dynamic is especially evident in educational systems marked by *tracking* (differentiated training pathways) and high institutional segregation. Sampermans et al. (2021) demonstrated that in Belgium (Flanders) and the Netherlands, an open classroom climate and school participation deepen civic knowledge inequalities between academic and vocational tracks. Students in the academic track (generally of higher SES) capitalize significantly better on these civic opportunities than their peers in the vocational track, worsening systemic inequality. Additionally, Hoskins et al. (2021) documented an isolated case of acceleration in Denmark, where school civic participation favored advantaged students more in their future voting intentions.

Methodological Strategy

To empirically test these competitive hypotheses, this work implements a quantitative research design structured into two independent yet theoretically articulated studies. Study 1 (Temporal Evolution in a Local Context – Chile): Examines the stability of civic gaps and school effects over two decades in a highly segregated educational system. Four databases with national representativeness are utilized: the Chilean samples from the IEA corresponding to CIVED 1999, ICCS 2009, and ICCS 2016 (all matching 8th–grade students), innovatively complemented with a Chilean sample of 10th–grade students assessed in the year 2019. The focus is to answer: to what extent are background inequalities observed regarding the intention to vote in future elections? and to what extent can school characteristics compensate for them?

The study 2 (Comparative International Analysis – ICCS 2016): Evaluates the same questions from a transnational perspective using the full international dataset of ICCS 2016 (the latest international version in which Chile participated with representative data). Unlike the first study, this shifts the focus toward cognitive performance, utilizing students' civic knowledge as the dependent variable across all participating countries.

Variables and Analytical Strategy

In both studies, background independent variables are operationalized through standardized composite indices of household Socioeconomic Status (SES)—incorporating parents' education and occupation, alongside cultural capital measured by the number of books in the home—while controlling for gender and migratory status. The school–level independent variables tested are: Open Classroom Climate (openness to discussion in the classroom), Participatory Activities within the institution, and the presence of a specific Civic Education Subject. Given that LSA data present a nested structure (students grouped within schools), the statistical analysis is conducted using Hierarchical Linear Modeling (HLM). The formal testing of the compensation hypothesis is executed by including cross–level interaction terms between the student's individual SES and each of the three school practice variables (SES×School_Practice).

Preliminary Results

The preliminary empirical analyses from both studies yield three central findings:

Persistence of Structural Inequalities: Significant background inequalities are observed in both the intention to vote in future elections (Study 1) and civic knowledge levels (Study 2). Family SES remains the most robust predictor of both outcomes, a pattern that stands unperturbed across the different years observed in Chile (1999–2019) and cross–nationally within ICCS 2016.

Relevance of the School: The analyzed school characteristics and practices are individually relevant for promoting civic outcomes. Classroom openness to discussion, fostering participatory activities, and the presence of civic education subjects display positive and statistically significant main effects on the average results of schools.

Absence of Compensation: Despite the direct positive impact of the school, the compensation hypothesis is virtually unconfirmed. The evidence consistently demonstrates that school practices do not produce a reduction or moderation of background inequalities. The interaction coefficients are statistically non–significant, indicating that the benefits of the school environment are distributed in a parallel manner without shortening the initial gap. This pattern of non–compensation is verified across different historical moments and cohorts in Chile (Study 1) and is robustly replicated when analyzing civic knowledge internationally with ICCS 2016 (Study 2).
The preliminary results force a critical reconsideration of the institutionalized pedagogical optimism embedded in citizenship education policies. The fact that school practices are positively associated with better individual civic outcomes, yet fail to alter baseline socioeconomic gaps, firmly places the evidence within the Maintenance Hypothesis(and potentially the Acceleration Hypothesis in highly segregated systems).

Keywords: School Principals, Accountability, School performance, Inequalities.

Screening non–invariance via alignment and orthogonal residuals. An illustration with students' background questionnaire across ICCS cycles

Diego Carrasco – Andres Sandoval Hernandez – Nurullah Eryilmaz – David Torres Irribarra

Introduction

International Large–Scale Assessments (ILSAs) such as TIMSS, PIRLS, and ICCS play a pivotal role in shaping educational research and policy worldwide. These assessments not only provide student achievement results in core domains such as mathematics, science, reading, and civic knowledge, but they also provide extensive background information on the conditions under which learning takes place. Through questionnaires administered to students, teachers, and school leaders, ILSAs collect rich contextual data that inform our understanding of the broader educational ecosystem.
The resulting background scales capture a wide variety of constructs—ranging from students' motivation, attitudes, and self–efficacy, instructional practices, and resources. These measures are indispensable for examining the mechanisms that explain differences in achievement, identifying equity gaps, and guiding policy interventions aimed at improving learning outcomes. In many cases, the background data provide the necessary explanatory power that achievement scores alone cannot deliver.
Despite their importance, a central methodological challenge remains: to what extent are these background scales comparable across countries and across time?
Cross–sectional comparability is difficult to achieve because cultural, linguistic, and contextual differences influence how items are interpreted and responded to (Sandoval et al., 2025a). For instance, a construct such as "school climate" may carry different connotations in different national settings, and translations of survey items may not capture nuances equally across languages. As a result, comparisons between countries can be contaminated by differences in interpretation rather than true differences in the underlying attribute.
Longitudinal comparability introduces an additional layer of complexity. Over successive cycles of ILSAs, background instruments are often revised to improve measurement, respond to emerging policy demands, or adapt to contextual changes. Such revisions may inadvertently alter the responses process, a phenomenon known in literature as scale drift (Bandalos, 2018). Even when items remain the same, changes in societal norms, educational practices, or respondent perceptions can affect how items are understood. Moreover, translation adjustments and shifts in survey administration modes (e.g., from paper to computer–based assessments) can further undermine comparability over time.

Measurement invariance methods are used to assess if measurement model among groups and across cycles can be held equal. The most popular approaches to assess if the same measurement model holds across groups of interest are differential item functioning (DIF), and factorial invariance (Thissen, 2024). The first asses if the expected responses over an item are similar among groups once attribute level differences among observations is accounted for. While the second, assess if the same measurement model can be held equal among the compared groups. When measurement invariance is not reached, researchers may search for partially invariant solutions to allow comparisons between countries, cycles, or subpopulation groups. That is response models where most parameters are held equal between groups, while other parameters are free to vary between groups. The generation of a partially invariant model is very time consuming, and requires substantial manual intervention (Svetina et al., 2020; Robitzsch & Lüdtke, 2023). In essence, building a partially invariant response model is a cumbersome process.

Recently, several studies in the large–scale assessment literature rely on the alignment optimization method (Sandoval et al., 2025a). The alignment methods (Muthén & Asparouhov, 2014; Asparouhov & Muthén, 2014; Asparouhov & Muthén, 2022) offer a friendly alternative to reach partially invariant solutions because the time investment on the side of the user is substantially reduced.

Studies using the International Civic and Citizenship Education Study (ICCS) have attempted to compare attitudes towards immigrants, political tolerance and political trust using alignment methods (Munck et al., 2018; Ziemes, 2024; Stals et al., 2024). Similar efforts have been done with the Trends in International Mathematics and Science Study (TIMSS) to generate comparable scores for scales from the background questionnaires (e.g., Glassow et al, 2021; Wurster, 2022; Yiğiter, 2024;) and with the Teaching and Learning International Survey (TALIS) (e.g., Fang et al., 2025), to name a few. A commonality across all these examples is the use of the alignment methods for confirmatory factor analysis (CFA), where responses to items are taken as continuous.

Although the previous studies have assumed CFA with continuous indicators can reach acceptable solutions to make comparison between countries, they did so because it was the only available option at the time (Munck et al., 2018, p33); and because they were not aware of any concerns of this modeling choice (Ziemes, 2024, p16). Nevertheless, fitting CFA models over polytomous items is not advisable, especially when items present 4 response options (Rhemtulla et al., 2012). Under this latter scenario, the CFA will underestimate factor loadings and parameter standard errors. Moreover, group comparison can get distorted due to the response modeling choice (Lubke et al., 2004). Treating ordinal responses as continuous measures in conditional models can lead to detects relationships that are not there, undetected relationships, and inversion effects (i.e., reordering of expected mean) (Liddell, et al. 2018). Thus, using CFA for categorical indicators should lead to fewer inferential errors regarding group comparison on scales from background questionnaires (Desa, 2014). In summary, the use of alignment method is widespread in large scale assessment, yet with a wrong model.

In the present study, we propose using confirmatory factor analysis for ordinal indicators (CFAord) (e.g., Bovaird & Koziol, 2012, Wang & Wang, 2020). We propose to first diagnose if measurement invariance is not reach with a model–based approach. That is, to compare between countries and cycles, if strict and scalar invariance is not satisfied with a CFA for ordinal indicators using appropriate model specification where thresholds are held equal (Wu & Estabrook, 2016; Svetina et al., 2020; Tse, et al, 2024). Then, for scenarios where there is evidence of lack of invariance, we propose to search for partially invariant solution using alignment methods while fitting CFA ord to reduce inferential errors for group comparisons between countries and across cycles. Finally, we propose a heuristic procedure based on the revised Angoff delta method (Magis & Facon, 2012) to search for groups that seems to sit closure on their measurement parameters.

Methods

Data

We used data from the International Civic and Citizenship Education Study from cycles 2009, 2016 and 2022. This study provides representative representative samples of 8th grade students, using a two–stage design, with intact classrooms (mean = 3866, min = 357, max = 7594). To illustrate the present approach, we treat each participating country and cycle as independent groups, thus allowing to assess measurement invariance between countries and between cycles simultaneously (see Munck et al., 2017; and Muthén & Asparouhov, 2014).

Compared groups. We include all participating countries and regions in ICCS 2009, 2016 and 2022, Participating countries and regions includes: Austria, Belgium (Flemish community), Bulgaria, Brazil, Switzerland, Chile, Colombia, Cyprus, Czechia, Denmark, Germany – North Rhine–Westphalia (benchmarking), Dominican Republic, Germany – Schleswig–Holstein (benchmarking), England (UK), Spain, Estonia, Finland, France, Greece, Guatemala, Hong Kong SAR, Croatia, Indonesia, Ireland, Italy, Republic of Korea, Liechtenstein, Lithuania, Luxembourg, Latvia, Mexico, Malta, Netherlands, Norway, New Zealand, Peru, Poland, Paraguay, Romania, Russian Federation, Serbia, Slovak Republic, Slovenia, Sweden, Thailand, and Chinese Taipei. Due to the varying participation between cycles, we count 86 country–cycle groups included in the study.

Dependent variables. We retrieved the common items from 8 different scales from ICCS across cycles 2009, 2016 and 2022 with five to six items. The item batteries include:

Table 1. Scales scores present in ICCS 2009. 2016, 2022 with 5 shared items or more

Number of items	Scale name
6	Students' perceptions of openness in classroom discussions
5	Students' perceptions of student–teacher relations at school
6	Students' trust in civic institutions
6	Students' endorsement of gender equality
5	Students' endorsement of equal rights for all ethnic groups in society
5	Students' sense of citizenship self–efficacy
5	Students' civic participation at school
5	Positive attitudes toward the influence of religion in society

Analytical strategy

Factorial invariance. To assess factorial invariance relying on confirmatory factor analysis for ordinal indicators we specify a series of different models following the recommendations of Wu & Estabrook (2016), Svetina et al. (2020) and Tse et al. (2024). We specify four model specification: configural, threshold, scalar, and strict models, and retrieved their fit indexes for comparisons. Rutkowski & Svetina (2017), relying on simulation studies for ordinal indicators with many groups (e.g., 10–20 groups) found that different model fit indexes cut off were not sensitive to deviations from invariance, and propose an RMSEA < .055 as a good indicative of fit for invariance. We relied on their proposal to assess invariance.

Alignment. We fit graded response models with alignment among the compared groups and retrieved the number of common parameters across groups. Muthén & Asparouhov (2017) suggest as soft rule of thumb aiming for aligned models with a proportion of .75 of invariant parameters. Only items shared across the three cycles are subject to analysis.

Orthogonal residual screening. We retrieve the partial credit model estimates of each compared group, and fit major axis line across each pair and calculate perpendicular distances of each threshold to this axis line. Moreover, w eproduce standard deviation for each pair (Magis & Facon, 2012), and

retained these to build a dissimilarity matrix among compared groups. Groups sharing a common measurement functioning should form low–dissimilarity blocks regardless of latent mean and variances. To illustrate the orthogonal residual screening approach, we focus only on the results of Students' trust in civic institutions, and Students' endorsement of equal rights for all ethnic groups in society, among Chile (2009, 2016) and Italy (2009, 2016, and 2022).

Preliminary results

Using the methods described above, we found that "Students' perceptions of student–teacher relations at school" and "Students' endorsement of equal rights for all ethnic groups in society" can be considered invariant using Rutkowski & Svetina (2017) criteria (RMSEA < 0.055). While the alignment soft rule of thumb for all the studied scales is not satisfied (75% of parameters can held equal) (see Table 2).

Table 2. Model based invariance indicators, and alignment proportion of invariant and varying model parameters.

				Invariance (RMSEA)			Alignment	
			N items	Strict	Scalar	Threshold	Invariant	Varying
1	S_OPDISC	Students' perceptions of openness in classroom discussions	6	0.086	0.084	0.080	0.43	0.57
2	S_STUTREL	Students' perceptions of student–teacher relations at school	5	0.060	**0.050**	0.056	0.57	0.43
3	S_INTRUST	Students' trust in civic institutions	6	0.096	0.101	0.123	0.45	0.55
4	S_GENEQL	Students' endorsement of gender equality	6	0.117	0.132	0.155	0.53	0.47
5	S_ETHRGHT	Students' endorsement of equal rights for all ethnic groups in society	6	0.066	**0.055**	0.064	0.55	0.45
6	S_CITEFF	Students' sense of citizenship self–efficacy	8	0.080	0.080	0.099	0.52	0.48
7	S_SCHPART	Students' civic participation at school	7	0.059	0.060	0.064	0.47	0.53
8	S_RELINF	Positive attitudes toward the influence of religion in society	6	0.068	0.066	0.079	0.54	0.46

Notes. Rutkowski & Svetina (2017) propose an RMSEA < .055 as a good indicative of fit for invariance models with ordinal indicators for many groups (e.g., 10–20 groups). Muthén & Asparouhov (2017) suggest as soft rule of thumb aiming for aligned models with a proportion of .75 of invariant parameters. Only items shared across the three cycles are subject to analysis.

The orthogonal residual screening illustration example results asserts that Italy and Chile cycles from 2016 are less linearly discrepant in their thresholds, while Italy 2009, 2022 with Chile 2009 result obtained item thresholds of less linear discrepancy, on both scales.

Conclusion and Discussion

The present study illustrate how difficult is to achieve measurement invariance across countries and cycles, despite common items. When invariance was assessed with confirmatory factor analysis for ordinal indicators across 86 country–cycle groups, only two of the eight scales met the RMSEA criterion of Rutkowski and Svetina (2017), and no aligned solution reached the 75% invariant–parameter benchmark of Muthén and Asparouhov (2017). The orthogonal residual screening procedure offers a practical complement: rather than a single global verdict, it locates subsets of groups whose thresholds are approximately exchangeable—as illustrated by Chile 2016 pairing with Italy 2016, and Italy 2009 and 2022 sitting closer to Chile 2009—showing that comparability is a relational property among specific groups rather than an all–or–nothing feature of a scale.

The screening procedure is heuristic: perpendicular distances summarize threshold discrepancy but do not quantify the consequences of non–invariance for latent mean rankings or trend estimates, and simulation work is needed to calibrate decision rules for when a block is comparable enough. Likewise, the RMSEA cutoff was derived from simulations with 10–20 groups, and its behavior with 86 groups remains an open question. Nevertheless, the procedure scales naturally to the full dissimilarity matrix, opening the possibility of reporting trends and cross–national comparisons within empirically comparable blocks—a defensible middle path between assuming comparability and abandoning comparison altogether.

Keywords: Invariance; Alignment; Background questionnaires; Citizenship

The Immigrant Paradox in Mathematics Self–Concept: A Cross–National Test Across 54 Education Systems Using TIMSS 2023

Maximiliano Romero – Rolf Strietholt – Ioulia Televantou

Introduction

Academic self–concept, students' perception of their own competence in a domain (Shavelson, Hubner, & Stanton, 1976), is distinct from achievement even though the two are reciprocally related (Marsh & Craven, 2006). Self–concept is also reference dependent. Students compare themselves with their peers, producing the big–fish–little–pond effect (Marsh, 1984), and across academic domains, as formalized in the internal/external frame of reference model (Marsh, 1986).

Beyond peers and domains, self–concept is also shaped by background. A recurrent finding, known as the immigrant paradox, is that immigrant students report more favorable outcomes than their circumstances would predict, in some cases even higher than non–immigrant peers once achievement and socioeconomic status are held constant (e.g., Coll & Marks, 2012; Marks, Ejesi, & Coll, 2014). Evidence for this pattern is more limited for mathematics self–concept, resting on a small set of studies (Areepattamannil & Freeman, 2008; Marsh, 2016). The most developed case, a robust first–generation advantage in Australian adolescents that persists after controls, comes from a single highly selective skilled–migration destination and motivates the SES–of–origin–country hypothesis (Basarkod et al., 2022), leaving open whether the pattern extends to other countries.

This leaves four gaps. First, generalizability: it is unknown whether the paradox is a general feature of self–concept formation or an artifact of migration regimes that differ across countries. Second, the developmental gap: most previous evidence concentrates on adolescents, leaving the early grades close to unexamined. Third, robustness: whether the advantage survives controls for absolute socioeconomic status is an indirect probe of the mechanisms proposed to explain it. Fourth, generational specificity: whether the pattern is concentrated in the first generation or extends to the second remains unsettled.

In the present study, we investigate the association between immigrant status and mathematics self–concept using international data. Specifically, we examine four associations:

1) The unadjusted association between immigrant background and mathematics self–concept.
2) The association after controlling for socioeconomic status and mathematics achievement.
3) The difference between first– and second–generation effects.
4) The variation in the association across countries.

Data Used

We use the Trends in International Mathematics and Science Study (TIMSS) 2023, Grade 4 (von Davier et al., 2024), covering 54 education systems and 336,892 students. Immigrant status is coded in three categories: first–generation immigrant, second–generation immigrant, and non–immigrant, based on students' and parents' country of birth. Mathematics self–concept is the TIMSS students' confidence in mathematics scale; mathematics achievement is the set of five plausible values; and socioeconomic status is the TIMSS home socioeconomic status index, entered at both the individual and school–mean level.

Methodology

We conduct an individual participant data (IPD) meta–analysis (Brunner et al., 2023). The analytic strategy has two stages. First, at the country level, three nested regression models are estimated separately within each education system: M1 regresses mathematics self–concept on immigrant status; M2 adds mathematics achievement; M3 adds individual and school–level socioeconomic status. Models use design–based regression with plausible values, sampling weights, jackknife replication, and multiple imputation of missing data. Statistical significance is approximated by $|\beta| > 2*SE$.

At the second stage, we conduct a random–effects meta–analysis of the country–specific coefficients to compute mean effect sizes and heterogeneity across countries. Heterogeneity is summarized through τ^2, I^2, and the prediction interval for country–specific effects.

Results and Discussion

The random–effects meta–analysis shows that the association between immigrant status and mathematics self–concept is small on average and highly heterogeneous across systems. For first–generation students, the unadjusted meta–analytic effect is negative and not distinguishable from zero (M1: $\beta = -0.022$, 95% CI [–0.078, 0.035]). Once achievement is added the effect turns positive (M2: $\beta = 0.041$, 95% CI [–0.005, 0.087]), and it remains at the same value, now barely excluding zero, once individual and school–level socioeconomic status are added (M3: $\beta = 0.041$, 95% CI [0.001, 0.080]). For second–generation students the unadjusted effect is negative and significant (M1: $\beta = -0.062$, 95% CI [–0.102, –0.021]), moves close to zero once achievement is added (M2: $\beta = 0.017$, 95% CI [–0.011, 0.046]), and is fully attenuated in the adjusted model (M3: $\beta = 0.007$, 95% CI [–0.022, 0.036]). Achievement, not socioeconomic status, accounts for the shift from a null or negative unadjusted association to a small positive one.

The central result concerns heterogeneity. Between–country variance remains substantial even in the fully adjusted model ($I^2 = 56.3\%$ for first–generation, 57.6% for second–generation M3 estimates), and the 95% prediction intervals span both negative and positive values at every specification (M3: [–0.158, 0.239] for first–generation, [–0.146, 0.160] for second–generation). A small average effect therefore conceals large cross–national differences, with the association between immigrant status and mathematics self–concept ranging from a clear disadvantage to a clear advantage across countries.

These results place cross–national heterogeneity at the center of the account. The immigrant paradox in mathematics self–concept is not a stable cross–national regularity. It is a context–dependent pattern whose average understates the range of what individual education systems display.

Conclusions

Testing the immigrant paradox in mathematics self–concept across 54 education systems in TIMSS 2023, in a broad cross–national test at Grade 4, we find a small and only marginally significant average advantage for first–generation immigrant students once achievement and socioeconomic status are held constant, no advantage for second–generation students, and substantial unexplained variation across countries. These mean effect sizes, however, rest on highly heterogeneous findings across countries: the prediction intervals cross zero at every model specification, indicating that the paradox is present in some education systems and absent, or reversed, in others. We conclude that the immigrant paradox in self–concept is best understood as a context–dependent pattern rather than a universal feature of immigrant students' academic adjustment, and that further work should focus on explaining this cross–national variation rather than establishing the existence of a single average effect.

Keywords: Immigrant paradox; Mathematics self–concept; Meta–analysis; TIMSS 2023

Identification of a Minimum Response Threshold in Large–Scale Standardized Assessments Using INVALSI Data

Luca Bungaro – Marta Desimoni

Introduction

Missing item responses are a common issue in psychological and educational assessment settings, where their occurrence is often unavoidable for a variety of reasons. The current literature can be broadly divided into two main approaches: the use of missing–data imputation methods (Dai, 2021; Rose et al., 2010; De Ayyala et al., 2001; Finch, 2008; Ludlow & O'Leary, 1999; Lord, 1974) and statistical modeling approaches that explicitly incorporate missing–response information within the estimation process (Glas & Pimentel, 2008; Holman & Glas, 2005; Moustaki & Knott, 2000; O'Muircheartaigh & Moustaki, 1999). However, a limitation of both approaches, conceptual rather than methodological in nature, is that they attempt to address the missing–data problem by recovering part of the missing information, either through imputation procedures or directly within the statistical model, without considering the implications that such recovery may have when reporting results at the individual level rather than solely at the aggregate level. Indeed, these methods can be applied equally to individuals who have completed 90% of the test and to those who have completed only 10%. It is intuitive that the amount of uncertainty associated with the resulting estimates differs substantially between these two situations. Nevertheless, such differences in measurement precision are difficult to communicate when an individual's competencies are ultimately summarized by a single score. One possible solution is to estimate ability levels and assign proficiency classifications only to respondents who have completed at least a minimum proportion of the test items. Such a strategy could substantially reduce the impact of measurement error and provide more precise and reliable estimates of individual achievement. The central aim, and innovative contribution, of the present study is therefore to develop, illustrate, and evaluate a statistically grounded methodology capable of identifying this minimum response threshold. Specifically, the proposed approach seeks to determine the minimum proportion of administered items required to obtain sufficiently accurate and reliable ability estimates for individual–level reporting.

Data used

The proposed methodology, which is described in greater detail in the following section, can be considered a "practical" approach, in the sense that the identification of a minimum response threshold depends directly on the characteristics of the data being analyzed, including test length, item difficulty, the distribution of students' responses, and the order in which items are presented. For this reason, the methodology was evaluated not only through a simulation study (also described in the Methodology section) but also using real data from the 2025 INVALSI Grade 13 Mathematics assessment. Specifically, the dichotomized response matrix of the entire sample was used, excluding respondents who had not viewed the complete test. The resulting dataset consisted of 1,939 examinees. However, the analyses were not conducted on the pooled sample, as they were performed separately for each test form developed by INVALSI for the 2025 main study. In particular, INVALSI developed 12 forms for each of the three major school categories (scientific lyceums, technical institutes, and vocational institutes/non–scientific lyceums), resulting in a total of 36 distinct test forms. Each form contained between 36 and 37 items. The 36 forms were being analyzed separately because, although they shared a subset of common items, the analysis of non–reached items depends directly on the position at which an item is administered within the test. This represents another substantial difference from the previously cited literature, which primarily focuses on the broader issue of missing values. In contrast, the present study is exclusively concerned with non–reached items, which INVALSI identifies and classifies differently from items that were reached by respondents but left unanswered. This distinction is particularly relevant in computer–based assessments, where different types of missing responses can be identified through process data. Within the INVALSI framework, missing–by–design responses (items not administered to a student), non–reached responses (items located after the last viewed item), and omitted responses (items viewed but left unanswered) are coded separately. While omitted responses are operationally treated as incorrect answers, non–reached responses are handled as missing data. This distinction is especially important because INVALSI provides individual–level feedback to students, making the interpretation of missing responses more consequential than in many traditional large–scale assessment contexts, where results are typically reported only at the aggregate level.

Methodology

Partially inspired by the comparative study of imputation methods conducted by Dai (2021), the proposed methodology is based on evaluating different proportions of non–reached items and calculating the corresponding Root Mean Square Error (RMSE) for each condition. The underlying reason for that is to identify a maximum RMSE threshold that can be considered acceptable, beyond which ability estimates can no longer be regarded as sufficiently reliable for reporting purposes. The study is divided into two phases. The first phase consists of a simulation study, whereas the second phase applies the proposed methodology to the real assessment data provided by INVALSI and described in the previous section. The simulation study was partially based on the framework proposed by Dai (2021). Specifically, several approaches for handling missing responses were compared: retaining missing responses as missing (MISS), treating missing responses as incorrect and assigning them a value of zero (IN), item mean imputation (IM), person mean imputation (PM), two–way imputation (TW), and response function imputation (RF). All simulations were conducted in R using the *mirt* package. This package was employed to generate item psychometric characteristics according to the Rasch model (Rasch, 1960), simulate examinees' response patterns, and estimate individual ability parameters. Missing–response imputation was performed using the *TestDataImputation* package, with all imputation methods implemented within the same computational framework to ensure comparability across conditions.

The simulation procedure was conducted as follows:

- Tests consisting of 40 items were simulated. Item difficulty parameters were generated according to the Rasch model (Rasch, 1960), assuming a standard normal distribution, N (0,1).
- The abilities of 1,000 students were also generated from the same standard normal distribution.
- Within each test, items were arranged according to the bell–shaped structure typically adopted in INVALSI Mathematics assessments: relatively easy items appeared at the beginning of the test, followed by items of medium and high difficulty, while the final section again contained relatively easier items.
- In each of the 100 simulation replications, all students were administered the same test.
- Non–reached items were simulated as follows. For each percentage level ranging from 10% to 90% in increments of 10% (10%, 20%, ..., 90%), the corresponding proportion of items at the end of the test was treated as non–reached. Furthermore, to reproduce the heterogeneity typically observed in real testing situations, the exact number of non–reached items was allowed to vary across students while maintaining the target percentage on average. For example, under the 10% condition, not all students had exactly 4 non–reached items; some had 2 or 3, whereas others had 5 or 6, thereby generating an overall average of approximately 10% non–reached items.
- After generating response data for each non–reached–item condition, the six missing–data handling methods were applied, and students' ability parameters were estimated. This entire procedure, including all previous steps, was repeated 100 times. Finally, the Root Mean Square Error (RMSE) was calculated for each imputation method and each percentage of non–reached items considered in the simulation.

$$RMSE = \sqrt{\frac{1}{N}\sum_{i=1}^{N}\left(\theta_i - \hat{\theta}_i\right)^2}\ , \tag{1}$$

where θ_i is the real ability (simulated) and $\hat{\theta}_i$ is the estimated one.

The results of this simulation study are presented in the following section. However, the methodological contribution of the present work extends beyond the simulation framework. As anticipated, a second phase of the study was conducted using real assessment data within a hybrid framework, combining observed data with simulation–based procedures. In this second phase, the comparison among imputation methods was no longer considered. Instead, only the MISS approach was retained, whereby non–reached items are treated as missing responses. This decision was motivated by two considerations. First, MISS is the official procedure adopted by INVALSI for handling non–reached items. Second, it is a widely used and well–supported approach in psychometric literature. Unlike the previous simulation study, neither the psychometric characteristics of the items nor the students' response patterns were simulated. Rather, both were obtained from the 2025 INVALSI Grade 13 Mathematics assessment. The only simulated component of this second approach concerned the generation of non–reached items, which was implemented through a procedure similar, but not identical, to that adopted in the previous phase. Specifically, rather than

removing predefined percentages of items, items were sequentially removed from the end of the test for each student until only one administered item remained. For example, for a student who completed a 36–item form, the procedure first treated the 36th item as non–reached, then the 35th and 36th items, then the 34th, 35th, and 36th items, and so forth. In this way, increasingly larger portions of the final section of the test were treated as non–reached. At each stage of this sequential item–removal procedure, students' ability parameters were re–estimated using the *TAM* package in R. The choice of this package was motivated by its ability to produce estimates that are fully consistent with those obtained within the INVALSI operational framework. The RMSE was subsequently computed for each condition. However, unlike in the simulation study, the true ability parameter (θ_i) was not available. Consequently, θ_i was defined as the ability estimate obtained from the complete test administered by INVALSI. For this reason, only students who had completed the entire test were included in the analyses. It is important to emphasize that the RMSE obtained in this second phase has a slightly different interpretation from that calculated in the simulation study. In the simulation framework, the RMSE reflected both the error introduced by non–reached items and the inherent measurement error associated with ability estimation. In contrast, the present approach assumes that the ability estimate derived from the complete test represents the true ability level. As a result, the RMSE isolates the specific component of error attributable to the presence of non–reached items. This procedure of sequential item removal, ability re–estimation, and RMSE calculation was performed separately for each of the 36 test forms included in the study. As an illustrative example, Table 1 presents the results obtained for Form 1 of the vocational school track.

Reached Items	36	35	34	33	32	31	30	29	28
RMSE	0.062	0.089	0.117	0.142	0.171	0.195	0.230	0.260	0.280
Reached Items	27	26	25	24	23	22	21	20	19
RMSE	0.297	0.315	0.341	0.358	0.371	0.403	0.425	0.446	0.483
Reached Items	18	17	16	15	14	13	12	11	10
RMSE	0.517	0.545	0.569	0.621	0.693	0.727	0.779	0.804	0.837
Reached Items	9	8	7	6	5	4	3	2	1
RMSE	0.889	0.952	0.994	0.996	1.024	1.066	1.238	1.410	1.618

Table 1. *RMSE values as a function of the maximum number of items reached for Form 1 of the vocational school track.*

As shown in Table 1, each value of reached items corresponds to a specific RMSE value, with the RMSE increasing as the number of reached items decreases. The next methodological step consisted of defining a maximum acceptable RMSE threshold in order to identify the corresponding minimum number of items that must be reached for reliable ability estimation. For the purposes of this study, three RMSE thresholds were considered: 0.4, 0.5, and 0.8 logits. The thresholds of 0.4 and 0.8 logits were derived directly from the proficiency–level framework adopted by INVALSI for Mathematics. Specifically, excluding the lowest and highest proficiency levels, which do not have a lower or upper boundary, adjacent proficiency levels are separated by intervals of 0.8 logits. Consequently, an RMSE threshold of 0.4 logits corresponds to half the width of a proficiency level, whereas a threshold of 0.8 logits corresponds to the full width of a proficiency level. The intermediate threshold of 0.5 logits was selected as a compromise between the other two values and corresponds to half of one logit unit. In the example reported in Table 1, achieving an RMSE strictly lower than 0.4 logits requires that a student has reached at least 23 items. The corresponding minimum number of reached items decreases to 19 when the RMSE threshold is set at 0.5 logits and to 12 when the threshold is set at 0.8 logits. In the following section, for each of the three RMSE thresholds, the median minimum number of reached items required across the 36 test forms is reported.

Results and discussion

The first important finding emerging from the simulation study concerns the performance of the six missing–data handling methods considered (Figure 1). Consistent with previous evidence reported in the literature on missing responses in general, the MISS approach exhibited one of the best estimation

performances when applied to non–reached items. The IM, IN, and RF methods also maintained satisfactory levels of performance across the conditions examined. In contrast, the TW and PM methods were characterized by substantially larger estimation errors, even when the proportion of non–reached items was relatively low.

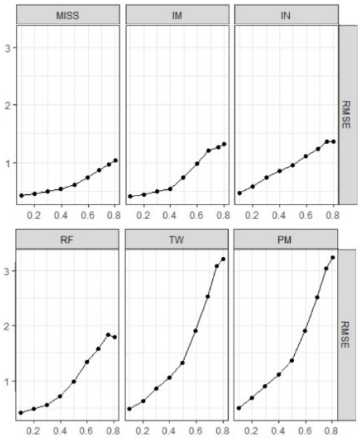


Figure 1. *RMSE trends across different proportions of non–reached items for the six missing–data handling methods.*

Overall, all methods exhibited an increase in RMSE as the proportion of non–reached items increased. However, although the relationship was consistently positive, the rate of increase was not constant across conditions. In most methods, the growth in RMSE accelerated as larger proportions of non–reached items were introduced. To examine this pattern more closely, Figure 2 reports the incremental changes in RMSE between consecutive non–reached–item conditions for the MISS approach only. Given its superior performance and its adoption within the INVALSI operational framework, the MISS method was selected for the subsequent analyses.

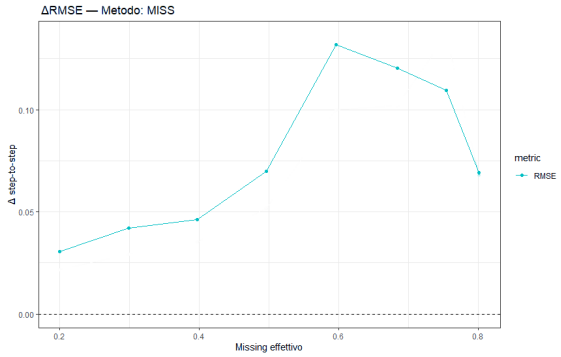


Figure 2. *RMSE increments across different proportions of non–reached items for the MISS approach.*

As shown in Figure 2, the largest increase in RMSE occurs between the 50% and 60% non–reached–item conditions. This finding suggests the presence of a critical point beyond which estimation accuracy deteriorates more rapidly. Based solely on the simulation study, a preliminary conclusion is that a plausible threshold may lie around 50% of items reached, corresponding to the point immediately preceding this sharp increase in RMSE. With regard to the analyses conducted on the real data from the 2025 INVALSI Grade 13 Mathematics assessment, the results were summarized into a single indicator representing the minimum number of items that must be reached under each of the three RMSE thresholds considered. These results are reported in Table 2.

Maximum Acceptable RMSE	0.4	0.5	0.8
Minimum Number of Items Reached	22	18	9
Minimum Number of Items Reached (%)	59%	49%	24%

Table 2. *Minimum number of items required to be reached as a function of the selected RMSE threshold. Grade 13 Mathematics, INVALSI 2025.*

As expected, the minimum number of required items decreases as the maximum acceptable RMSE threshold increases. These findings become particularly informative when interpreted alongside the results of the simulation study. Specifically, if attention is shifted from the absolute number of items to the proportion of items reached relative to the maximum test length (37 items), the thresholds associated with RMSE values of 0.4 and 0.5 correspond closely to the 50% benchmark identified in the simulation analyses. In contrast, the 0.8–logit threshold is considerably more permissive, allowing students to reach only 24% of the test while still being considered eligible for ability estimation. However, this greater flexibility comes at the cost of substantially larger estimation error.

Conclusions

In conclusion, both the simulation study and the application to the 2025 INVALSI Grade 13 Mathematics assessment demonstrated that the proposed methodology can effectively exploit students' response data to identify a minimum item–reached threshold for producing more precise and reliable ability estimates. The findings obtained from the two complementary approaches converge toward the same conclusion: the methodology provides a clear and empirically grounded answer to the question of how many items must be reached before an individual ability estimate can be considered sufficiently reliable. At the same time, the methodology offers a high degree of flexibility, allowing decision makers to select the maximum acceptable level of measurement error according to the objectives and requirements of a particular assessment context. Consequently, different minimum item–reached thresholds can be established depending on the desired balance between estimation precision and the inclusion of a larger number of respondents in the reporting process.

Keywords: Simulation; RMSE; Missing Item Responses; New Methodology

SESSION 16. EDUCATIONAL INEQUALITIES, LEARNING OUTCOMES AND OPPORTUNITIES IN CHANGING CONTEXTS

ORGANIZERS: UNIVERSITY OF BERGAMO – UNIVERSITY OF MILAN "LA STATALE"

COORDINATORS: JURGENA MIYFIU – GIULIANA PARENTE

6TH NOVEMBER: 8.30 A.M. – 10.30 A.M. {AULA B0.1 – RESEARCH 18}

Using AI without reducing the person to a dataset: study methods using AI and AI–assisted learning assessment to tackle educational inequalities

Letizia Ferri

Ethical framework and reference framework

In the age of artificial intelligence (AI), the educational risk is not merely technical but anthropological: the loss of wisdom and the interruption of the 'passing of the baton', that is, the transmission of knowledge and methodology from one generation to the next. The encyclical *Magnifica Humanitas* (Leo XIV, 2026) calls for the wisdom needed to interpret technology and warns against stifling young people's desire to ask questions in the face of machines that make human thought seem superfluous (no. 140); it identifies the school as a place where one learns to seek and love the truth, built on shared time and trustworthy relationships; it calls for a 'fast from AI' and for the elimination of inequalities in access to education (no. 147), without reducing the person to a mere data point. This perspective engages with the philosophy of information: Floridi (2014, chap. 7) describes 'enveloping', that is, the tendency to re–ontologise the world to make it 'tailored' to AI – a warning against reshaping schools and their practices to suit the needs of machines, but rather to govern technology. On the regulatory and ethical front, Benanti (2018) proposes 'algorethics' – the incorporation of ethical criteria into algorithms – whilst emphasising that 'intelligent machines' do not replace human wisdom and decision–making. Added to this are the MIM Guidelines (2025), which adopt equity as a principle – AI must not exacerbate inequalities – and promote an anthropocentric AI, and the European Commission's ethical guidelines (2022), which call for a critical and ethical approach, in accordance with the GDPR. The premise of this paper is therefore a sustainable and anthropocentric use of AI, dedicated to the transmission of knowledge and methodology.

The programmes: FSL (School–to–Work Training) and preparatory courses

The first core element of this paper is a framework of workshop programmes, co–designed by schools and the University of Insubria, dedicated to study methods using AI. The School–Work Training programme 'Studying with Artificial Intelligence' (formerly PCTO) does not 'teach how to use AI', but teaches students to study more effectively by making use of it, whilst keeping critical thinking, source evaluation and individual responsibility at the centre. It is structured around three dimensions – curricular (core elements of the disciplines: text analysis, argumentation, writing), experiential ('learning by doing') and guidance–based (reflection on one's own method and choices) – and unfolds in stages: a critical introduction to the functioning and limitations of AI systems; a methodology workshop (mind maps, summaries and assisted rephrasing, with a critical comparison between one's own work and AI–generated content); critical review (identifying errors, biases and 'hallucinations', verifying and citing sources); inclusion, with AI as a tool for support and personalisation for students with specific learning difficulties (SLD) and special educational needs (SEN); project work culminating in a final digital artefact (podcast, e–book, presentation). Assessment is both process–based and outcome–based (rubrics, observation sheets, logbook, digital portfolio). Alongside the school curriculum are university preparatory courses, designed for the transition from secondary school to university – a stage typically characterised by cumulative disadvantage: they prepare future students to develop a sustainable study method using AI, with the same critical and inclusive approach, so that the knowledge acquired is not lost during the transition. The common aim is precisely that set out in the encyclical: to ensure that AI supports, rather than replaces, the passing on of knowledge, whilst reducing the digital divide that separates those who use AI methodically from those who are merely passive recipients of it.

Data validation using AI as a means of assessing learning

The second key area concerns the way in which these learning pathways are monitored and assessed. The qualitative data they generate – reflections, logbooks, portfolios, exit tickets – are rich but costly to analyse, and this hinders their effective use. An AI–assisted analysis procedure is therefore proposed as a tool for

assessing learning, both in terms of process and outcome: multiple AI systems independently generate their own inductive coding scheme and apply it to the texts; the results are triangulated; each system's scheme is subjected to the others (cross–validation), separating the 'scheme effect' from the 'coder effect'; reliability is estimated using percentage agreement and Cohen's k (Cohen, 1960; Landis & Koch, 1977), and a human arbitration process resolves any disagreements. The aim is to obtain reliable evidence rapidly that can be used immediately to review training and support learning, whilst also identifying differential effects across student profiles (the equity lens). The same analytical approach is currently being used, in articles in press, to assess the work and learning of trainee teachers (60 CFU programmes). In line with the ethical framework, AI accelerates the process and makes the data more accessible, but does not replace evaluative judgement: this avoids reducing learning to mere data (Benanti, 2018; Leone XIV, 2026, no. 147).

Conclusions

The proposal integrates inclusive study method programmes using AI – FSL and former preparatory courses – aimed at preserving knowledge and passing on the torch, alongside a rapid and reliable method of data analysis for assessing learning. The overall approach is geared towards countering educational inequalities in the AI era, addressing the digital divide, transitions and inclusion (DSA/BES), within an explicit ethical framework (Leone XIV, 2026; Floridi, 2014; Benanti, 2018; MIM, 2025; European Commission, 2022). Work is ongoing: plans include the first pilot testing of the learning pathways with the measurement of differential outcomes by profile, the development of a blind human 'gold standard' for the assisted assessment procedure, and a comprehensive report on the multi–model design. Limitations include: the pilot scale, linguistic and domain–specific nature, reproducibility linked to model versions, and ethical and data protection safeguards.

Keywords: Study methods; Critical thinking; FSL (formazione Scuola Lavoro); AI–assisted learning assessment; AI ethics

Understanding tertiary education mobility among migrant – origin students in Italy

Maria Berrittella – Francesco Salomone Marino

The increasing globalization of higher education has contributed to a substantial rise in student mobility, creating new opportunities for educational advancement while also raising concerns about unequal access to these opportunities. Among the groups most affected by these dynamics are students with migrant backgrounds, whose educational trajectories are shaped by the interplay of migration–related experiences, family resources, academic achievement, and institutional contexts. Although mobility for tertiary education is often considered a key mechanism for enhancing educational and occupational prospects, relatively little is known about the factors influencing mobility decisions among migrant–origin students and the extent to which these choices reproduce or reduce existing inequalities. Thus, the main aim of this paper is to examine the determinants of tertiary education mobility among migrant–origin students in Italy. In particular, the study investigates how individual characteristics, family background, school performance, and school–level contexts influence students' decisions to pursue higher education either within or outside their region of origin. To address these objectives, we apply multinomial logistic regression models to longitudinal data drawn from the MOBYSU.IT database (2020), an integrated dataset combining information from the National Archive of Schools, the database of the National Evaluation Institute for the School System, and the National Archive of University Students. This unique data source allows us to follow migrant–origin students from upper secondary education into tertiary education and to analyze mobility choices within a comprehensive life–course perspective.

The findings indicate that tertiary education mobility is a highly selective process influenced by both socioeconomic resources and academic performance. Factors such as parental educational attainment, household economic resources, educational aspirations, and individual academic skills significantly increase the likelihood of undertaking mobility for higher education. Students with stronger academic records and higher levels of cultural capital are more likely to enroll in universities outside their local area, suggesting that educational mobility is often associated with the accumulation of advantages that facilitate access to broader educational opportunities.

At the same time, the analysis reveals substantial heterogeneity among migrant–origin students. Those from less advantaged socioeconomic backgrounds face greater constraints in accessing mobility opportunities, even when they demonstrate high academic achievement. Financial limitations, reduced access to information and guidance, and lower levels of family support may hinder their ability to translate educational potential into mobility decisions. These findings suggest that educational mobility is not solely determined by merit, but is also shaped by unequal access to economic and social resources.

School performance and school–level characteristics emerge as additional factors influencing mobility choices. Students attending high–performing schools are more likely to pursue tertiary education outside their region of residence, benefiting from stronger academic preparation, better guidance services, and more ambitious educational environments. Conversely, schools serving disadvantaged populations may provide fewer opportunities and resources to support students in navigating higher education pathways, thereby limiting mobility prospects. These results highlight the important role of educational institutions in either reinforcing or mitigating existing inequalities among migrant–origin students.

Overall, the results demonstrate that tertiary education mobility among migrant–origin students is shaped by a complex interaction between structural opportunities and individual agency. Family resources, academic achievement, school contexts, and migration–related factors jointly influence mobility decisions and contribute to the reproduction of educational inequalities. The findings suggest the need for targeted policies aimed at reducing economic and informational barriers, strengthening guidance and support services, and promoting equitable access to mobility opportunities in higher education. Such interventions are essential to ensure that educational mobility functions as a mechanism of social inclusion and upward mobility rather than as a process that reproduces existing social and educational disparities between native and migrant–origin students.

The increasing globalization of higher education has contributed to a substantial rise in student mobility, creating new opportunities for educational advancement while also raising concerns about unequal access to these opportunities. Among the groups most affected by these dynamics are students with migrant backgrounds, whose educational trajectories are shaped by the interplay of migration–related experiences, family resources, academic achievement, and institutional contexts. Although mobility for tertiary education is often considered a key mechanism for enhancing educational and occupational prospects, relatively little is known about the factors influencing mobility decisions among migrant–origin students and the extent to which these choices reproduce or reduce existing inequalities.

The main aim of this paper is to examine the determinants of tertiary education mobility among migrant–origin students in Italy. In particular, the study investigates how individual characteristics, family background, school performance, and school–level contexts influence students' decisions to pursue higher education either within or outside their region of origin. To address these objectives, we apply multinomial logistic regression models to longitudinal data drawn from the MOBYSU.IT database (2020), an integrated dataset combining information from the National Archive of Schools, the database of the National Evaluation Institute for the School System, and the National Archive of University Students. This unique data source allows us to follow students from upper secondary education into higher education and to analyze mobility choices within a comprehensive life–course perspective.

The findings indicate that tertiary education mobility is a highly selective process influenced by both socioeconomic resources and academic performance. Factors such as parental educational attainment, household economic resources, educational aspirations, and individual academic skills significantly increase the likelihood of undertaking mobility for higher education. Students with stronger academic records and higher levels of cultural capital are more likely to enroll in universities outside their local area, suggesting that educational mobility is often associated with the accumulation of advantages that facilitate access to broader educational opportunities.

At the same time, the analysis reveals substantial heterogeneity among migrant–origin students. Those from less advantaged socioeconomic backgrounds face greater constraints in accessing mobility opportunities, even when they demonstrate high academic achievement. Financial limitations, reduced access to information and guidance, and lower levels of family support may hinder their ability to translate educational potential into mobility decisions. These findings suggest that educational mobility is not solely determined by merit but is also shaped by unequal access to economic and social resources.

School performance and school–level characteristics emerge as additional factors influencing mobility choices. Students attending high–performing schools are more likely to pursue tertiary education outside their region of residence, benefiting from stronger academic preparation, better guidance services, and more ambitious educational environments. Conversely, schools serving disadvantaged populations may

provide fewer opportunities and resources to support students in navigating higher education pathways, thereby limiting mobility prospects. These results highlight the important role of educational institutions in either reinforcing or mitigating existing inequalities among migrant–origin students.
The study also suggests that mobility may generate important benefits for migrant students. Those who undertake educational mobility tend to access higher–quality educational opportunities, develop broader professional and social networks, and strengthen their intercultural and adaptive competencies. Nevertheless, mobility experiences may also involve challenges related to financial pressures, social integration, feelings of belonging, and identity negotiation, particularly for students who lack adequate support mechanisms.
Overall, the results demonstrate that tertiary education mobility among migrant–origin students is shaped by a complex interaction between structural opportunities and individual agency. Family resources, academic achievement, school contexts, and migration–related factors jointly influence mobility decisions and contribute to the reproduction of educational inequalities. The findings underscore the need for targeted policies aimed at reducing economic and informational barriers, strengthening guidance and support services, and promoting equitable access to mobility opportunities in higher education. Such interventions are essential to ensure that educational mobility functions as a mechanism of social inclusion and upward mobility rather than as a process that reproduces existing social and educational disparities.

Keywords: High school; INVALSI test scores; Migrant–Origin Students; Student mobility; Tertiary education

Teaching Practices, Learning Outcomes and Educational Inequalities in Lower Secondary Schools: Evidence from INVALSI Data

Jurgena Myftiu – Giorgia Casalone

Educational inequalities remain a central challenge for contemporary education systems. Students' learning outcomes are still strongly associated with socioeconomic background, family resources, migration status and territorial context, raising concerns about the capacity of schools to guarantee equal learning opportunities (Coleman et al., 1966; Breen & Jonsson, 2005; Sirin, 2005; OECD, 2019). While a large body of research has focused on individual and family determinants of achievement, increasing attention has been devoted to school–level and classroom–level factors that may either mitigate or reinforce educational disparities.
Among these factors, teachers' instructional practices, assessment strategies and use of digital learning environments represent key mechanisms through which schools can shape students' opportunities to learn. Teacher collaboration and active instructional practices are commonly interpreted as central components of professional learning communities and collective school capacity (Stoll et al., 2006; Vangrieken et al., 2015). International evidence also shows that collaborative professionalism may support school improvement by fostering shared responsibility, peer exchange and coherent pedagogical strategies (Hargreaves & O'Connor, 2018; OECD, 2020). Empirical studies further suggest that teacher collaboration can be positively associated with student achievement, especially when it is embedded in instructional teams and shared pedagogical work (Goddard et al., 2007; Ronfeldt et al., 2015).
This paper investigates whether different teaching practices are associated with students' learning outcomes and with the reduction of educational inequalities in lower secondary education. The empirical analysis focuses on Grade 8 students in Italy and combines INVALSI standardised test scores in Italian and Mathematics with information from the 2022/2023 INVALSI teacher questionnaires. The focus on Grade 8 is particularly relevant because it corresponds to the end of lower secondary education, a crucial transition point in students' educational trajectories.
The study contributes to the literature in three main ways. First, it shifts the focus from individual teacher characteristics to the quality and structure of classroom–level learning opportunities. Second, it integrates information on both Italian and Mathematics, allowing the analysis to compare whether similar teaching practices operate differently across disciplinary domains. Third, it explicitly links teaching practices to educational inequalities by examining whether they moderate the relationship between students' socioeconomic background and achievement.
The INVALSI teacher questionnaires provide a rich set of indicators on instructional practices in Italian and Mathematics (INVALSI, 2023a; INVALSI, 2023b). In both subjects, teachers report the frequency with which

they use information and communication technologies, including computers, interactive whiteboards, tablets, e–learning platforms and educational software. They also provide information on active and collaborative teaching practices, such as sharing materials through cloud platforms, promoting classroom discussion, using peer–to–peer and cooperative learning, asking students to present work in different formats, adopting flipped classroom activities, and organising laboratory–based learning. Assessment practices are also observed, including the use of group work evaluation, structured and semi–structured tests, digital platforms for assessment, authentic tasks and multimedia products. These dimensions are particularly relevant because assessment and feedback are widely recognised as central components of learning processes and instructional effectiveness (Black & Wiliam, 1998; Hattie & Timperley, 2007).

The questionnaires also include information on how teachers use students' mistakes: as a basis for revising methodological choices, as a resource for approaching content differently, as an opportunity for discussion among students, or as an occasion for reflection with colleagues. This dimension is especially relevant for the study of learning processes and critical thinking, because it captures whether error is treated as a purely corrective event or as a pedagogical resource for reflection, discussion and improvement.

In addition to these common dimensions, the questionnaires include subject–specific indicators. For Mathematics, the analysis will consider opportunity–to–learn measures related to the main INVALSI domains: numbers, relations and functions, space and figures, and data and probability. Particular attention will be paid to activities involving problem solving, modelling, argumentation, conjectures, use of data and interpretation of graphs and tables. For Italian, the analysis will consider indicators related to lexical competence, reading comprehension, interpretation of narrative–literary texts, interpretation of expository and argumentative texts, discussion, argumentation, metacognition, and grammar teaching. These subject–specific indicators make it possible to capture not only the intensity of teaching practices, but also the extent to which students are exposed to cognitively demanding and reflective learning activities.

The empirical strategy proceeds in two steps. First, synthetic indicators of teaching practices will be constructed. The main specification will rely on exploratory factor analysis, using polychoric correlations given the ordinal nature of most questionnaire items. Factor analysis is appropriate because it allows the identification of latent dimensions underlying a set of correlated observed variables and is widely used for the construction of theoretically meaningful constructs from survey data (Fabrigar & Wegener, 2011). The analysis will identify dimensions such as active and collaborative learning practices, digital integration, formative and authentic assessment, reflective use of errors, and subject–specific opportunity to learn. The adequacy of the factor solution will be assessed using standard diagnostics, including the Kaiser–Meyer–Olkin statistic, Bartlett's test of sphericity, factor loadings and internal consistency measures.

As a robustness check, the same dimensions will also be summarised through the Adjusted Mazziotta–Pareto Index, a non–compensatory composite indicator that is particularly useful when the aim is to measure multidimensional phenomena while penalising unbalanced profiles across components (Mazziotta & Pareto, 2016). After normalising the indicators and assigning a positive polarity to items where higher values correspond to richer learning opportunities, the AMPI can be defined as:

$$AMPI_i^- = M_{z_i} - S_{z_i} \cdot CV_{z_i}$$

where M_{z_i} is the mean of the normalised indicators for unit i, S_{z_i} is their standard deviation, and CV_{z_i} is the coefficient of variation. The negative sign is used because the phenomenon is interpreted positively: higher values indicate more frequent or richer teaching practices, while the penalty term reduces the score when the profile is highly unbalanced across dimensions.

In the second step, the resulting indicators will be linked to students' standardised achievement scores using multilevel models. This approach is appropriate because INVALSI data are hierarchically structured: students are nested within classes and classes within schools. Multilevel modelling makes it possible to separate individual, classroom and school–level sources of variation, while accounting for the dependence among observations within the same educational context (Raudenbush & Bryk, 2002; Snijders & Bosker, 2012). The baseline model can be written as:

$$Y_{icsm} = \beta_0 + \beta_1 TP_{csm} + \beta_2 ESCS_{ics} + \beta_3 (TP_{csm} \times ESCS_{ics}) + \beta 4 \mathrm{X}_{ics} + \beta 5 \mathrm{Z}_{csm} + \beta 6 \mathrm{W}_s + u_s + v_{cs} + \varepsilon_{icsm}$$

where Y_{icsm} is the standardised test score of student i in class c, school s, and subject m, with m referring to Italian or Mathematics. TP_{csm} represents the teaching–practice indicator constructed through factor analysis or AMPI. $ESCS_{ics}$ captures students' socioeconomic and cultural background. The interaction term

$TP_{csm} \times ESCS_{ics}$ is the key parameter for the analysis of inequalities: it tests whether the association between socioeconomic background and achievement is weaker or stronger in classrooms characterised by richer teaching practices. X_{ics} includes student–level controls such as gender, migration background and regularity of school progression; Z_{csm} includes teacher or classroom characteristics where available; and W_s includes school–level and territorial controls. Finally, u_s and v_{cs} are school and class level random effects, while ε_{icsm} is the individual error term.

The empirical analysis will estimate separate models for Italian and Mathematics, as well as pooled models including subject fixed effects. In addition to continuous achievement scores, the study will also consider the probability of being a low performer as a secondary outcome, using multilevel logistic models. This second outcome is particularly relevant from a policy perspective because it allows the analysis to identify whether specific teaching practices are associated with a lower risk of educational vulnerability.

Three research questions guide the study. First, which latent dimensions of teaching practices emerge from the INVALSI Grade 8 teacher questionnaires? Second, are active, collaborative, digital and formative teaching practices associated with higher learning outcomes in Italian and Mathematics? Third, do these practices reduce educational inequalities by weakening the relationship between students' socioeconomic background and achievement?

The expected contribution is both empirical and policy–oriented. If classrooms characterised by active learning, formative assessment, reflective use of errors and cognitively demanding disciplinary practices show weaker socioeconomic gradients, this would suggest that teaching practices can play a role in mitigating inequality of opportunity. At the same time, the study will interpret the results cautiously, since teaching practices are not randomly assigned across classrooms and schools. The analysis is therefore designed to provide robust evidence on associations rather than strict causal estimates.

Overall, the paper contributes to the debate on educational inequalities by showing how classroom–level learning opportunities interact with students' background characteristics. It also provides evidence on the organisational and pedagogical mechanisms through which schools may promote more equitable learning outcomes. In line with the focus of the session on learning outcomes and opportunities in changing contexts, the study highlights the importance of teaching practices not only for average achievement, but also for the distribution of educational opportunities across students.

Keywords: Teaching practices; Learning outcomes; Multilevel models; Composite indicators Teaching practices; Learning outcomes; Multilevel models; Composite indicators

A simplified socio–cultural index for measuring educational inequalities in learning outcomes

Clelia Clascella

Introduction and research motivation

Educational poverty and inequality of opportunity remain central challenges for contemporary education systems, particularly in contexts characterised by persistent social, economic, and territorial disparities. (OECD, 2023; Coleman et al., 1966) Within this debate, a crucial analytical question concerns how to conceptualise and measure the role of students' socio–economic and socio–cultural background in shaping learning outcomes. While it is widely acknowledged that educational outcomes are strongly associated with family background, the magnitude, interpretation, and policy implications of this relationship largely depend on how socio–economic status (SES) is operationalised. (Sirin, 2005; Marks et al., 2006)

The literature has produced a variety of SES measures, often leading to inconsistent or incomparable results across studies and contexts. (Avvisati, 2020; OECD, 2017) Divergences in the choice of indicators (particularly regarding parental education, occupation, and family wealth) have made it difficult to draw cumulative evidence on educational inequality. For example, while traditional SES indices adopt a multidimensional approach, the inclusion of heterogeneous and sometimes weakly comparable indicators of family wealth has raised concerns about validity, comparability, and measurement error, especially in cross–national research. (Avvisati, 2020; Engzell, 2021)

This study contributes to the debate on the measurement of educational poverty, understood here as a manifestation of unequal educational opportunities, by proposing and validating a simplified socio–cultural index (SCINDEX). (Reardon et al., 2013; Cascella, 2025) The index provides a parsimonious yet effective

measure of students' socio–cultural background, enabling more robust and comparable analyses of inequality in educational outcomes.

By focusing on measurement issues, this research directly addresses the first thematic area of the conference call, while also contributing to broader discussions on cumulative disadvantage and the role of structural inequalities in shaping educational trajectories.

Conceptual Framework: Educational Poverty and Socio–Cultural Background

Educational poverty can be conceptualised as a multidimensional condition reflecting unequal opportunities to acquire cognitive, cultural, and social resources necessary for academic success. (Reardon et al., 2013) Rather than focusing solely on achievement gaps, this perspective emphasises the structural constraints that limit students' capabilities, including family background, migration status, and territorial context.

Within this framework, SES serves as a key proxy for unequal opportunities. Extensive empirical evidence shows that students from low SES backgrounds systematically underperform compared to their more advantaged peers, often entering secondary education with substantial skill gaps. (Sirin, 2005; OECD, 2023) These disparities are further compounded by interactions with other dimensions of disadvantage, such as gender and ethnicity, reinforcing cumulative inequalities over time.

However, the operationalisation of SES remains contested. (Avvisati, 2020) While there is general agreement on its multidimensional nature (typically including parental education, occupation, and income or wealth) the relative importance and measurement of these components are debated. In particular, recent research has questioned the validity of commonly used indicators of family wealth (e.g., household possessions), highlighting problems of cross–cultural comparability and bias in self–reported data. (Engzell, 2021). Building on these critiques, the present study adopts a socio–cultural perspective on educational poverty, emphasising the role of cultural and occupational resources transmitted through families. This approach aligns with evidence suggesting that cultural capital (proxied by parental education and occupation) may play a more stable and comparable role than material resources in explaining educational inequalities. (Marks et al., 2006; Cascella, 2025)

Data and Methodology

The empirical analysis is based on data collected by the Italian National Institute for the Evaluation of the Education System (INVALSI), focusing on representative samples of students in Grade 5 (primary school) and Grade 10 (upper secondary school). (Campodifiori et al., 2010) These data are particularly suitable for the study because they combine standardised test scores in mathematics with detailed background information on students and their families.

The analytical strategy consists of three main steps:

1. Assessment of SES indicators: The contribution of each component of the widely used ESCS index (Economic, Social, and Cultural Status) is examined through multilevel regression models, accounting for the hierarchical structure of the data (students nested within schools).
2. Construction of SCINDEX: A simplified typological index is developed, combining the highest parental education and occupation into a single measure of socio–cultural background.
3. Validation and comparison: The predictive validity of SCINDEX is compared with that of ESCS in explaining students' mathematics achievement.

The multilevel modelling approach allows for disentangling individual and contextual effects, thus aligning with the conference's emphasis on integrating micro– and macro–level determinants of educational inequality.

Key Findings

The Relative Importance of SES Components

Results indicate that parental education and occupation are strongly and consistently associated with students' learning outcomes. (Sirin, 2005; Marks et al., 2006) Higher levels of parental education and occupational status correspond to significantly higher achievement in mathematics.

In contrast, indicators of family wealth (measured through home possessions) contribute relatively little to explaining variation in achievement once education and occupation are accounted for. (Engzell, 2021; Avvisati, 2020) This finding supports the hypothesis that traditional proxies for material resources may no longer adequately capture socio–economic conditions, particularly in contemporary and comparative contexts.

The Effectiveness of the SCINDEX

The SCINDEX, despite being based on fewer indicators, demonstrates predictive validity comparable to that of the ESCS index. (Cascella, 2025) Specifically, the strength of the association between socio–cultural

background and academic achievement remains essentially unchanged when using SCINDEX instead of ESCS.

This result suggests that a parsimonious measure focusing on parental education and occupation can effectively capture the socio–cultural dimension of educational inequality without relying on potentially problematic wealth indicators.

Robustness Across Educational Stages

The findings are consistent across both primary and secondary education levels, indicating that the simplified index is robust across different stages of schooling. (Reardon et al., 2013; OECD, 2023) This highlights the persistence of socio–cultural inequalities throughout educational trajectories, reinforcing the notion of cumulative disadvantage.

Research implications

Implications for the Measurement of Educational Poverty

The study contributes to the measurement of educational poverty in several ways.

First, it demonstrates that simplified and theoretically grounded indicators can capture essential dimensions of inequality without compromising explanatory power. This is particularly important for research contexts with limited data availability or resources, where comprehensive SES measures may be difficult to implement.

Second, the removal of weak or unstable indicators (such as certain proxies of family wealth) enhances comparability across contexts and over time. This is crucial for advancing cumulative knowledge and for conducting reliable comparative analyses, particularly in international research.

Third, the use of a typological index such as SCINDEX facilitates interpretation and policy communication, as it categorises students into meaningful socio–cultural groups rather than relying on abstract continuous scores.

Overall, the findings support a shift towards more parsimonious, transparent, and context–sensitive approaches to measuring educational inequality.

Policy Relevance and Implications

The results have important implications for educational policy and the design of interventions aimed at reducing educational poverty.

By highlighting the central role of socio–cultural factors, the study underscores the importance of policies that enhance cultural capital and educational support within families, particularly among disadvantaged groups. (Coleman et al., 1966; OECD, 2023) Interventions may include parental education programmes, early childhood education initiatives, and school–based support targeting students from less advantaged backgrounds.

Furthermore, the persistence of SES–related disparities across different educational stages suggests that policies should adopt a life–course perspective, addressing inequality early and preventing its accumulation over time.

Finally, the availability of a simplified and cost–effective index such as SCINDEX can support evidence–based policymaking, particularly in contexts where data collection resources are limited. By facilitating the measurement and monitoring of educational inequality, it provides a valuable tool for evaluating the effectiveness of policy interventions.

Conclusion

This study contributes to ongoing debates on educational poverty and inequality of opportunity by proposing a simplified yet robust measure of socio–cultural background. (Avvisati, 2020; OECD, 2017) The SCINDEX demonstrates that it is possible to capture the core dimensions of SES (and their impact on learning outcomes) without relying on complex or potentially unreliable indicators of family wealth.

By improving comparability, reducing measurement costs, and maintaining explanatory power, the proposed index represents a valuable tool for both researchers and policymakers. More broadly, the findings reinforce the importance of addressing socio–cultural inequalities as a key determinant of educational poverty and of designing policies that promote equal opportunities across diverse social and territorial contexts.

Keywords: Educational poverty; Socio–cultural background; Equality of opportunity; SES measurement; Multilevel analysis; Territorial inequalities; Italy

Detecting Low–Effort Response Behaviour in Digital Competence Assessment Using INVALSI DigComp 2025 Data

Chiara Rita Cunsolo – Mariagiulia Matteucci – Stefania Mignani

Introduction

Large–scale educational assessments are often administered in low–stakes contexts, where students' results have limited direct consequences for them. For this reason, students may not always be fully engaged while answering the test items. When this happens, the assumption that item responses reflect the latent trait being measured may be violated, because responses may also be influenced by students' level of test–taking engagement. If this aspect is ignored, the interpretation of test results and ability estimates may be affected.

This issue is particularly relevant for computer–based assessments, where response accuracy can be analysed jointly with response times. Response times provide additional information on the response process and can help distinguish between engaged and low–effort response behaviour.

In this study, we investigate disengaged response behaviour in the INVALSI DigComp2025 assessment, which measures students' digital competences in Italy (Vuorikari, Kluzer, and Punie, 2022). We apply a latent response model without omissions, inspired by the framework proposed by Ulitzsch, von Davier, and Pohl (2020), to distinguish between engaged and disengaged response processes at the item–by–student level. The aim of the study is to evaluate whether response behaviour in the DigComp assessment shows evidence of disengagement and to examine how accounting for such behaviour can support a more valid interpretation of students' digital competence.

Data

The analysis uses data from the INVALSI DigComp 2025 assessment, designed to measure students' digital competences in Italy. The assessment was administered in a computer–based format and included item responses together with response–time information. The DigComp assessment consisted of 22 test forms, with the number of items varying across forms from 23 to 28.

Since the latent response model was estimated separately for each form, the analysis was conducted at the form level. Due to the complexity of the model and the limited information on disengaged behaviour in some forms, successful estimation was achieved for 15 out of the 22 forms. The results reported in this study therefore refer to the subset of forms for which the model converged successfully.

The present application focuses on a model without omissions, so disengaged response behaviour is identified through the joint pattern of item responses and response times rather than through omitted responses.

In addition to the cognitive test data, the INVALSI DigComp 2025 assessment also includes questionnaire data. These data collect students' self–reported information on aspects related to digital competence, digital practices, and attitudes towards the use of digital technologies. In the present study, the questionnaire data are not analysed directly, but they represent a relevant direction for future extensions of the model.

Methodology

The analysis is based on a latent response model for detecting low–effort response behaviour. The model is a simplified version of the hierarchical latent mixture model proposed by Ulitzsch, von Davier and Pohl (2020), which builds on mixture modelling approaches for response accuracy and response times (Wang and Xu, 2015). The model is adapted to the present application by excluding the omission process. In this setting, each item–by–student interaction is assumed to arise from one of two latent response states: an engaged state and a disengaged state.

Let Δ_{ij} denote the latent engagement indicator for student i on item j, with $\Delta_{ij} = 1$ indicating engaged response behaviour and $\Delta_{ij} = 0$ indicating disengaged response behaviour. The probability of being engaged is modelled as:

$$P(\Delta_{ij} = 1) = \frac{\exp(\phi_i - \iota_j)}{1 + \exp(\phi_i - \iota_j)},$$

where ϕ_i represents the student's tendency to engage with the test (engagement of the student i), and ι_j represents the engagement difficulty of item j.

Conditional on the engaged state, the probability of a correct response follows a Rasch model:

$$P(Y_{ij} = 1 | \Delta_{ij} = 1) = \frac{\exp(\theta_i - b_j)}{1 + \exp(\theta_i - b_j)},$$

where θ_i denotes the student's latent ability and b_j denotes item difficulty. Thus, engaged responses are assumed to provide information about the target competence.
In contrast, conditional on the disengaged state, responses are assumed to be unrelated to the student's ability. The probability of a correct response is therefore governed by a common parameter c:

$$P(Y_{ij} = 1 | \Delta_{ij} = 0) = c.$$

Response times are used to distinguish between the two latent states. Let t_{ij} denote the response time for student i on item j. In the engaged state, log–response times are modelled as a function of student speed and item time intensity (van der Linden, 2007):

$$\log(t_{ij}) \mid \Delta_{ij} = 1 \sim N(\beta_j - \tau_i , \sigma_E^2) ,$$

where τ_i denotes student speed and β_j denotes item time intensity. In the disengaged state, response times are assumed to follow a separate distribution:

$$\log(t_{ij}) \mid \Delta_{ij} = 0 \sim N(\beta_D , \sigma_D^2) .$$

The model was estimated separately for each test form within a Bayesian framework, using Hamiltonian Monte Carlo as implemented in Stan/RStan (Stan Development Team, 2025). Four parallel chains were run with 4,000 iterations each, with the first half discarded as warm–up. Convergence diagnostics were inspected to determine which forms could be retained for substantive interpretation.

Results

The results are based on the 15 retained test forms, out of the 22 administered forms, for which posterior estimation was considered acceptable. Across these forms, the model identified a non–negligible proportion of item–by–student interactions as belonging to the disengaged response process.
The estimated disengagement rate averaged 7.0% across forms, with form–level estimates ranging from 6.0% to 8.1%. This suggests that low–effort response behaviour was present in the INVALSI DigComp 2025 assessment, although its prevalence varied moderately across test forms.
Response–time results support the interpretation of the disengaged component as a low–effort response process. The observed mean response time was 38.1 *seconds*, whereas the estimated mean response time for disengaged responses was 16.9 *seconds*. Thus, responses classified as disengaged were faster than the average observed response. This pattern is consistent with the idea that students in the disengaged state did not fully engage in the solution process but tended to provide faster responses.
At the same time, the results indicate that disengaged responses should not be interpreted as purely random guessing. The estimated low–effort response parameter c averaged 0.38 across the retained forms, with form–level estimates ranging from 0.31 to 0.49. Since the chance level for four–option items is 0.25, these values indicate that students classified as disengaged answered correctly more often than expected under completely random guessing. This suggests that the disengaged state may also include rapid partially informed answers, rather than only random guesses.
The model also revealed a positive association between students' engagement and their latent digital competence. Across the 15 retained forms, the average correlation between engagement and ability was 0.35. This indicates that students with higher estimated digital competence also tended to show higher engagement with the test.
Overall, these findings show that observed performance in the DigComp assessment reflects both students' digital competence and their level of test–taking engagement. Accounting for disengaged response behaviour is therefore important for improving the validity of interpretations based on the assessment results. In particular, the model helps distinguish responses that are more likely to reflect engaged solution behaviour from responses that are faster and less informative about the target competence.

Conclusions

This study shows that low–effort response behaviour is present in the INVALSI DigComp 2025 assessment and can be detected by jointly modelling item responses and response times. Across the retained test forms, the estimated disengagement rate was not negligible. Moreover, disengaged responses were much faster than the average observed response time. These results support the interpretation of disengagement as a distinct response process that should be considered when interpreting students' digital competence.

The findings also show that disengaged behaviour should not be interpreted only as random guessing. The estimated low–effort response parameter was consistently above the chance level expected for four–option items. This means that some disengaged responses may still contain partial information about the item, even if students did not fully engage in the solution process.

Accounting for test–taking engagement is important for the validity of the results. Observed performance may reflect both students' competence and their engagement with the test. Separating these two components can therefore lead to a more accurate interpretation of assessment results. It can also help identify forms, items, or response patterns that are more affected by low–effort behaviour.

Future work will aim to extend this approach in two directions. First, the same approach could be applied to the DigComp questionnaire data to investigate non–cognitive response behaviour, such as inattentive or low–effort answering in self–report measures. Second, future analyses could explore how to integrate the cognitive response model and the questionnaire response model within a common framework. This would make it possible to study cognitive and non–cognitive response behaviour jointly and to obtain a more complete picture of students' engagement in the DigComp assessment.

Keywords: Digital competence; Engagement; Latent response model; Low–effort response behaviour

SESSION 6 DATA LITERACY AND DATA USE IN SCHOOLS: EMERGING PRACTICES, CHALLENGES, AND PROFESSIONAL CULTURES

ORGANIZERS: FONDAZIONE BRUNO KESSLER – FONDAZIONE PER LA SCUOLA

COORDINATORS: DAVIDE AZZOLINI – VERONICA MOBILIO

6TH NOVEMBER: 11.00 A.M. – 1.00 P.M. {AULA MAGNA – RESEARCH 19}

From Data Decomposition to Curricular Design: "Deconstructing" INVALSI Outcomes as a Governance Methodology for Equity and Innovation. A Three–Year Case Study (2024–2026)

Andrea Meucci – Silvia Bennati – Fleana Albertoni

Introduction: Data Literacy Facing Systemic Challenges

This paper analyzes the data–use model adopted by the I.C. "Dante Alighieri" of Aulla over the 2024–2026 triennium. The theoretical framework is grounded in Data Literacy not as a mere bureaucratic requirement, but as an indispensable professional posture to steer school autonomy toward equity [INVALSI, 2026]. While the national educational landscape still highlights a persistent difficulty in fully recovering pre–pandemic learning levels [INVALSI, 2025], the school's action–research aims to demonstrate how the arithmetic mean of a school building or institute can act as an "anesthetizing" factor, obscuring deep internal disparities and macro–structural variability.

The core hypothesis is that only through a meticulous "deconstruction" of data — namely, a multidimensional disaggregation based on between–class variability and systematic cross–referencing with background factors (ESCS, gender, origin) — is it possible to promptly identify areas of "internal educational poverty" and, simultaneously, quantify teaching effectiveness in terms of value–added. This approach allowed the school to anticipate the implementation of new national educational policies [Ministero dell'Istruzione e del Merito, 2025], effectively transforming standardized assessment into a "compass" for designing a resilient, future–ready skills curriculum.

Data Sources

The research relies on a longitudinalized and individualized institutional dataset that includes:

- **INVALSI Outcomes (2024–2026):** A detailed, sequential analysis of cohort data (focusing on Grade 5 for primary school and Grade 8 for lower secondary school) in Italian, Mathematics, and English (Reading/Listening) [I.C. "Dante Alighieri" Aulla, 2025c].
- **Family Context Variables and Equity Indicators:** Application of the ESCS index for micro–mapping the correlation between socio–economic background and educational success, combined with the analysis of between–class and within–class variability parameters provided in official reports [INVALSI, 2025].
- **Strategic Documents:** Targets from the Self–Evaluation Report (RAV), process objectives from the School Improvement Plan (PdM), and the newly drafted 2025–26 Institutional Curriculum [I.C. "Dante Alighieri" Aulla, 2025a, 2026].

Methodology: The Deconstruction Process and Value–Added Estimation

The proposed methodology converts statistical analysis into pedagogical action through three sequential steps:

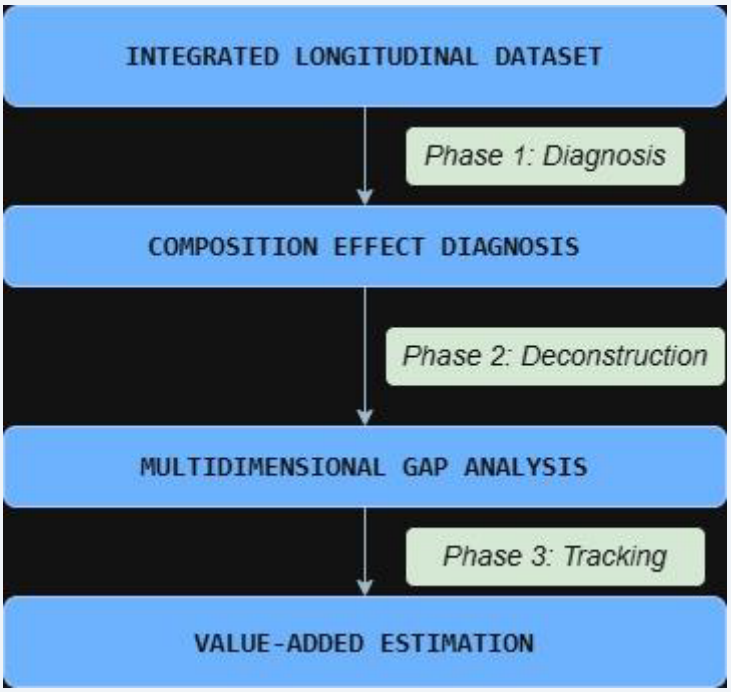


Figure 1: *Action–research methodological framework for equity–oriented governance: 1. Diagno–*
sis of the composition effect by isolating between–class macro–variability; 2. Multidimensional Gap Analysis by cross–referencing individual student performances with the ESCS index; 3. Estimation of value–added for tracking the actual institutional school–effect.

- **Phase 1: Isolating Variability and Diagnosing the Composition Effect.** The Institute isolates instances where between–class variability significantly exceeds territorial benchmarks. A pilot case is represented by Grade 5 Mathematics in 2025, where internal variability reached 54.5% compared to the national average of 12.7%, revealing the existence of "two parallel schools" under the same roof. Internal investigations traced the root of this gap to a sharp "composition effect" stemming from family organizational and scheduling choices. These choices concentrated a significantly higher ESCS index and a lower incidence of non–native Italian speakers in the full–time section compared to the standard–time section [I.C. "Dante Alighieri" Aulla, 2025c].
- **Phase 2: Multidimensional Gap Analysis and Background Control.** Results are further "deconstructed" by crossing the ESCS index with individual student performances. Grade 8 data highlight a stark polarization: students from low–ESCS backgrounds meet proficiency targets in Italian in only 52.6% of cases, compared to 93.3% in the high–ESCS bracket. For foreign languages, data mapping follows the Common European Framework of Reference (CEFR) to isolate specific oral reception deficits [Council of Europe, 2018].
- **Phase 3: Estimating the School–Effect and Tracking Latent Potential.** To verify whether performance depends on a mere composition effect or on the specific pedagogical efficacy of the class council, the methodology integrates the reading of the value–added (school–effect) provided by INVALSI. This approach enables the mapping of "Top Performers" (Level 5), monitoring how initially disadvantaged or highly heterogeneous environments still succeeded in yielding significant excellence shares.

Results and Discussion: From Diagnosis to the Improvement Plan (PdM)

The strategic and disaggregated use of data drove targeted interventions that profoundly restructured the school's organizational and instructional asset:

- **Mitigating and Reducing Implicit School Dropout.** Thanks to the early monitoring of students stuck at proficiency Levels 1 and 2 and the activation of custom mentoring tracks, the school's implicit dropout rate dropped from a historical 25.4% baseline down to 8,9% in 2025. This trajectory successfully preempts and consolidates the national targets set by the school's PNRR milestones for the 2026 deadline [INVALSI, 2025].

- **Inter–class Organizational Flexibility.** Scientific evidence of high between–class variability prompted the breaking of rigid, traditional group–class structures. This led to the institutionalization of open classes for macro–subject areas, parallel ability grouping, and the implementation of instructional "learning pauses" dedicated to active recovery and enhancement [I.C. "Dante Alighieri" Aulla, 2025a].
- **Early Vertical Curriculum Recalibration.** Specific linguistic weaknesses detected in Grade 2 (where an excellent reading comprehension score contrasted with a low 50% accuracy in structural morphosyntactic exercises) guided a full revision of the primary school curriculum, moving language structure reinforcement up to the first two–year cycle [I.C. "Dante Alighieri" Aulla, 2025c].

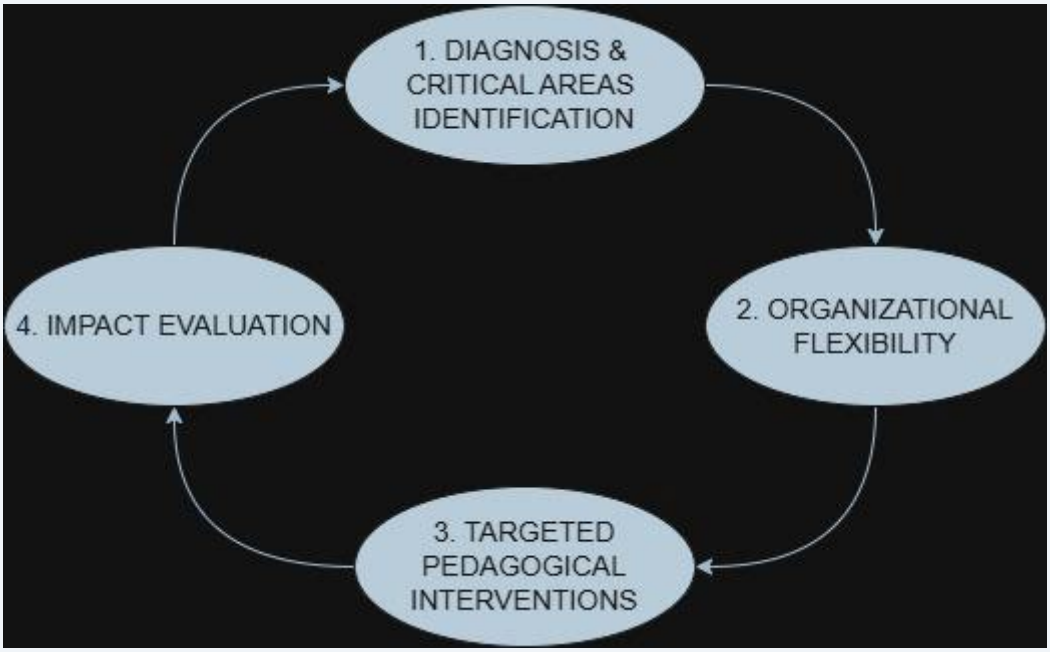


Figure 2: *Data–driven school governance and continuous improvement cycle: 1. Diagnosis and identification of critical areas; 2. Organizational flexibility (open classes across sections); 3. Targeted interventions (mentoring projects); 4. Impact evaluation on institutional outcomes.*

- **Fostering Absolute Excellence.** The deployment of differentiated learning paths based on value–added data fostered an overall expansion of talent in Grade 8 assessments, with the overall proportion of top–tier students (Level 5) in Italian rising from 8.9% to 11.36% year–over–year.

Future Outlook: The New Curriculum Facing National Guidelines

The final phase of this research projects data utilization onto the institute's new pedagogical and regulatory architecture. The 2025–26 Institutional Curriculum, renewed to incorporate the latest national reforms, frames INVALSI metrics in a predictive key, systematically embedding the European digital competence framework [Vuorikari et al., 2022].

- **Anticipatory Governance:** The school mapped and structured "Digital Citizenship" goals starting right from Kindergarten, proactively aligning with the assessment frameworks of the new national tests scheduled for the subsequent educational tiers [INVALSI, 2025].
- **Computer Science and STEM as Vectors for Equity:** The implementation of computer science as a vertical educational axis responds directly to the logical–mathematical weaknesses that emerged from the macro–domain analysis of the triennium, ensuring equal laboratory access across every single ESCS quartile [I.C. "Dante Alighieri" Aulla, 2026].

Conclusions

This study demonstrates that the Data Literacy of school staff represents the fundamental prerequisite for transforming standardized assessments from an exercise in mere external accountability into a genuine lever for educational equity and systemic resilience. The "deconstruction" model and the value–added analysis implemented by I.C. "Dante Alighieri" prove that it is entirely possible to defuse the impact of socio–economic backgrounds and manage internal school variability, turning student heterogeneity into a constructive starting point for personalized, inclusive, and internationally aligned education.

Keywords: Data Literacy in Education; School Governance and Equity; Data Deconstruction; School Value–Added

Reading Between the Lines: Data Literacy Trajectories in Service Learning Self–Assessment in Italian Schools (2022–2025)

Patrizia Lotti – Lorenza Orlandini

Introduction

The expansion of data available to schools is not limited to standardised assessments or digital platforms: it also encompasses the organisational self–assessment instruments that schools themselves generate over time. The Rubric "Maturità di impiego dell'approccio pedagogico del Service Learning" – an adaptation for the Italian K–12 context of Andrew Furco's Self–Assessment Rubric for the Institutionalization of Service–Learning in Higher Education (Furco, 1999, 2002a, 2002b) – represents a particularly informative case in this respect: it is a data infrastructure produced by schools about themselves, administered annually since 2022, yielding both a quantitative positioning on a closed scale (three developmental stages, 21 components across 5 dimensions) and a free–text qualitative narrative for each component. The contribution proposes to read this longitudinal corpus not only as a source of evidence on the institutionalisation of Service Learning (SL), but as an exemplary case of organisational data literacy: schools' capacity to interpret with growing precision the data they themselves produce about their own practice. This is a notion of data literacy less tied to individual technical competence and closer to research traditions that define data use as a social and situated process, rooted in professional cultures and organisational structures before individual capacity (Coburn & Turner, 2011). In this perspective, the Rubric is not an external measurement instrument but an internal reflective device: its repeated administration activates the synergy between institutional self–assessment and organisational development that research on schools as learning organisations identifies as a condition for continuous improvement (Admiraal, Schenke & de Jong, 2019).

Data

The corpus consists of four administrations of the Rubric over the period 2022–2025. Each year, Italian state and accredited private schools receive access to an online platform and complete the instrument on a voluntary basis. The Rubric is structured around 21 components distributed across 5 thematic dimensions (Philosophy and Mission, Teachers and Innovation, Student Engagement, Community Partnership, Institutional Support). For each component, schools indicate their developmental stage on a three–level closed scale – T1 (building critical mass), T2 (building quality), T3 (institutional sustainability) – and produce a free–text narrative documenting their concrete practices. At the close of each data collection, schools receive an individual report comparing their positioning against the national median: this feedback loop is the condition that makes each administration an act of service research rather than mere data collection. The sample has grown progressively: 59 schools in 2022, 97 in 2023, 120 in 2024, and 195 in 2025, for a total of 471 responses. A sub–set of 24 schools is present in three or more administrations and constitutes the longitudinal panel on which trajectory analysis is focused. The Rubric's components were localised to the Italian school context through the integration of the innovation dimensions developed within the "Avanguardie Educative" movement, while the three stages (T1–T2–T3) reflect the original developmental model elaborated by Furco (1999, 2002a, 2002b) and widely taken up in the international literature: from the phase in which initial experiences emerge through individual teachers' initiative (T1), to institutional diffusion and formalisation (T2), through to full integration in the curriculum and the school's evaluative instruments (T3), following a linear and cumulative sequential progression (Kecskes & Muyllaert, 1997).

Methodology

The analysis is guided by four research questions: (1) Which school actors interpret the data produced by the Rubric, and to what ends? (2) Which beliefs and attitudes shape the quality of this interpretation? (3) How does this quality vary according to membership of SL networks? (4) Which system–level strategies support a meaningful and situated use of data over time?

The analysis combines longitudinal tracking of quantitative positionings with thematic coding of the narratives following the principles of Thematic Analysis (Braun & Clarke, 2006), conducted across successive administrations on a corpus coded line by line. The methodological centrepiece is the triangulation procedure between the closed scale and the narrative: each thematic code was assigned an

implicit level (T1/T2/T3–equivalent) based on its semantics, then compared with the stage declared on the scale. Applied to 4,098 scale–narrative pairs, this procedure operationalises the interpretive quality of self–assessment through three categories: congruence, underreporting (the narrative describes more advanced practices than those declared on the scale), and overreporting (the scale declares a more advanced stage than the narrative documents). For the longitudinal sub–set, a mean institutionalisation score was also computed per year (T1=1, T2=2, T3=3), classifying trajectories as progression, stability, or regression.

Results and Discussion

The scale–narrative triangulation shows that 49% of pairs are congruent, 33% constitute underreporting, and 18% overreporting. Overreporting is not uniformly distributed: it reaches 32% on teaching methodologies (2.5b), 30% on integration in the PTOF (the school's statutory three–year educational plan) (1.3b), and 30% on partner engagement (4.1b), while falling below 5% on the evaluation of the innovation process (5.3b). This profile directly addresses the question of which beliefs shape data use practices: overreporting clusters where the gap between identity narrative and organisational practice is structurally widest, that is, where the school has built a narrative of excellence that struggles to translate into stable routines.

The most direct evidence of learning in data use emerges from the comparison of successive administrations for the same school. An illustrative case: a secondary school in Central Italy positioned itself at T3 on component 2.3b (interdisciplinarity) in 2024, while its narrative described activities that had "entirely fallen back into extracurricular activities to support the achievement of process objectives linked to the school's PTOF strategies" – content coded as T1–equivalent. In 2025, the same school corrected the scale to T1, keeping the narrative consistent. This episode of spontaneous self–correction, also detectable at the aggregate level (schools with more participation tend to reduce overreporting over time), constitutes direct evidence that repeated engagement with a self–assessment instrument produces progressive calibration of interpretive capacity. Not a side effect, but the outcome of a feedback mechanism.

A second indicator of data literacy emerges on the underreporting side: the thematic code relating to "tacit knowledge" of SL (schools that practise the approach without formally recognising it as such) grows from 2 occurrences in 2022 to 12–15 in 2025. A secondary school in Lombardy articulates this explicitly: "The SL dimension is not always recognised even when it is enacted. A purely lexical and labelling problem" (2024). This growth does not signal a deterioration of practice but a diagnostic maturation: by 2025, schools have developed more refined conceptual tools – through internal reflection on the results of previous administrations – for identifying the gap between what they do and what they recognise themselves as doing.

Regarding variation by school type, SL networks act as selective accelerators of data quality: networked schools show systematically higher T3 rates, with differences of up to +26 percentage points on SL definition (1.1b) and +18 points on the staff/working group component (5.1b). The structural exception is interdisciplinarity (2.3b), where even networked schools remain at T1 in 53% of cases: network membership improves the interpretive quality of data but does not remove the organisational constraints – timetabling, teacher turnover, difficulties in co–teaching arrangements – that pedagogical practice would require to advance. Networked schools are further distinguished by the quality of the language with which they describe SL: more frequent use of standardised identity formulas, explicit references to network–based professional development, and descriptions of community partnership in terms of co–design rather than simple collaboration.

On the role of artificial intelligence: in the 2025 administration, only 2 of 195 participating schools make explicit reference to AI tools in their narratives. Among schools practising SL, the data culture remains largely narrative and reflective, not mediated by algorithmic instruments.

The following cases were selected according to a criterion of maximum variation with respect to the phenomenon of self–assessment calibration: two schools showing documented spontaneous correction between successive administrations, two counter–types illustrating alternative patterns (high–quality stability; stagnation through staff turnover), and a fifth case selected for its analytically distinct value — the only school in the entire corpus to overcome the structural bottleneck of interdisciplinarity. All belong to the longitudinal panel (three or more administrations).

Case 1 – Spontaneous correction on a critical component *(upper secondary school, Lazio).* In 2024 the school positioned itself at T3 on component 2.3b (interdisciplinarity), while the corresponding narrative described exclusively extracurricular activities — T1–equivalent content. In 2025 it spontaneously corrected the scale to T1, keeping the narrative consistent. The episode is the most explicit case of progressive self–assessment

calibration in the entire corpus: the school has learned to distinguish between declaring timetable flexibility and enacting it structurally.

Case 2 – Diagnostic maturation without formal self–correction *(high school, Lombardy).* The school does not modify its scale positioning, but its 2024 narratives articulate with growing precision the gap between practice and recognition: «The SL dimension is not always recognised even when it is enacted. A purely lexical and labelling problem.» The phenomenon is one of conscious underreporting: diagnostic capacity grows independently of any change in score – an indicator of data literacy that the closed scale alone cannot capture.

Case 3 – High–quality stability *(primary and lower secondary school, Lombardy).* Present in three administrations with a stable mean score around 2.5, the school replicates virtually identical narrative descriptions between 2024 and 2025. The institutional SL project formalised by a School Board resolution in January 2021 is described in almost the same words across both years: «The gradual transition to the institutional SL project 'Mi riguarda' led, in January 2021, to a School Board resolution formalising the project.» This narrative invariance is not an absence of reflection: it signals that the practice has become routinised and no longer requires evolutionary narration. The finding is interpretable as consolidated institutionalisation — the most mature form of organisational data literacy, in which the instrument no longer generates surprises because the school already knows itself.

Case 4 – Stagnation through staff turnover *(primary and lower secondary school, Friuli–Venezia Giulia).* The only school present across all four administrations with a non–progressive profile. The mean score oscillates between 1.88 and 2.0: the narratives reveal an asymmetry — component 5.3b (evaluation of the innovation process) remains stable at T3, while component 2.1b (teacher knowledge and awareness) stays at T1. The mechanism is made explicit in the 2024 narrative: «In previous years, various training courses were organised and several projects were presented. Unfortunately, most of the teachers who implemented them are no longer part of the school's teaching staff.» The cycle of loss and reconstruction of professional capital prevents cumulative calibration: the school can evaluate its own process (stable T3 on 5.3b) but cannot transmit the acquired competence to new teachers.

Case 5 — Overcoming the structural bottleneck of interdisciplinarity *(primary and lower secondary school, Sicily).* An analytically distinct case: the only school in the longitudinal corpus to move from T1 to T3 on component 2.3b in a single administration (Δ=+0.52 between 2024 and 2025). The narrative explicitly attributes the leap to access to PNRR funds (Italy's National Recovery and Resilience Plan), which made possible a structural reorganisation of the school timetable. The link between financial leverage and institutionalisation is not incidental: reaching T3 on the interdisciplinarity component requires SL to be embedded in the regular curriculum through co–teaching arrangements and cross–disciplinary co–planning — a condition that implies a structural modification of the timetable difficult to sustain with ordinary resources. The case documents that this threshold cannot be crossed through progressive maturation of organisational culture alone, but requires an external lever acting on the structural conditions that block it.

Conclusions

The data analysed suggest that the institutionalisation of SL does not follow a linear trajectory but proceeds through successive steps and consolidation thresholds. The T1→T2 transition appears to occur, in many cases, when a school engages with the Rubric, with training materials made available by territorial networks and the Avanguardie Educative movement. More complex is the T2→T3 passage, which constitutes the main critical point in the institutionalisation process: reaching full institutional sustainability requires not simply broadening SL experiences or increasing their diffusion within the school, but the consolidation of specific organisational conditions — stability of the teaching staff, distributed leadership, recognised working groups, and the formalisation of choices in deliberative acts, in the PTOF or other institutional documents. The component concerning interdisciplinarity and school time organisation (2.3b) reveals a further difficulty: in this case, professional culture growth and network participation alone are insufficient to produce stable advancement. Interdisciplinary planning requires shared working time, co–teaching arrangements, coordination spaces and, in some cases, additional financial resources. This is therefore a threshold that directly involves school organisation and the contractual conditions of teaching work, prior to and independently of professional beliefs or competences alone.

In this perspective, the Rubric can function as a lever for data literacy only if embedded in a feedback mechanism that allows the school to compare its own positioning against an external reference and to return to reading it the following year. Diagnostic capacity – distinguishing between naming a practice and enacting it in a structurally integrated way – is not a precondition of the instrument but a measurable

outcome of its use over time. This suggests that the discussion on data literacy in schools should encompass, beyond data availability and quality, the explicit design of feedback cycles that transform data collection into organisational learning, and the recognition that some SL institutionalisation thresholds require interventions on structural conditions prior to, and independently of, professional culture.

Keywords: Self–assessment data; Data literacy; Longitudinal monitoring; Service–learning

Beyond Compliance: Internal Roles for Evidence–Informed Peer Review and Organisational Learning in Vocational Education and Training

Concetta Fonzo – Laura Evangelista

Introduction

Education and training providers are increasingly expected to make systematic use of data, evidence and feedback to support improvement, accountability and professional learning. In vocational education and training (VET), this expectation is particularly demanding because quality depends on several interdependent processes: curriculum design, work–based learning, guidance, inclusion, assessment, relations with enterprises, documentation, monitoring of outcomes and resource management. These processes generate information, but information alone does not produce improvement. Data become useful only when organisations are able to interpret them, discuss them collectively and transform them into sustainable decisions.

This contribution addresses the organisational conditions that allow Peer Review in VET to become an evidence–informed practice rather than a procedural requirement. The paper proposes a framework of internal roles that can support the collection, interpretation and use of evidence throughout the Peer Review cycle. The central argument is that Peer Review can promote organisational learning when it is not limited to external feedback or compliance with quality standards, but is embedded in distributed responsibilities, professional dialogue and improvement planning.

The theoretical framework combines three strands of literature. The first concerns quality assurance in VET, with specific reference to the EQAVET approach, which defines quality as a cyclical process of planning, implementation, evaluation and review. From this perspective, quality assurance is not simply the verification of external standards; it requires organisations to read their own processes, collect evidence, discuss results and convert feedback into improvement actions. The second strand concerns organisational learning. Following Argyris and Schön (1978) and Senge (1990), organisations learn when they transform experience into shared knowledge, question established routines and modify practices, decision criteria and coordination mechanisms. The third strand concerns collaborative and evidence–informed evaluation in education. The literature highlights the tension between accountability–oriented models, which may produce compliance and documentation, and collaborative models, which strengthen professional reflection, peer learning, feedback and distributed leadership.

Within this framework, the research gap concerns the weak explicit definition of the internal roles that make Peer Review practically effective. Manuals and procedures often describe phases, documents and operational steps, but they pay less attention to the people who prepare the process, collect and interpret evidence, facilitate dialogue, manage the visit, read the recommendations and translate them into actions. The paper therefore asks: which internal functions must be activated for Peer Review to support organisational learning and evidence–informed improvement in VET centres?

Data used

The study is based on three sources of data and evidence. First, it uses institutional and policy documents on quality assurance, EQAVET, Peer Review and VET quality management, including European recommendations and guidance documents that frame quality as a continuous improvement cycle. Second, it uses scientific literature on organisational learning, collaborative evaluation, peer learning, school and VET quality assurance, and the use of evidence in educational organisations. Third, it draws on semi–structured interviews with qualified informants who have direct experience of Peer Review, quality assurance or evaluation processes in educational and training contexts.

The empirical material does not aim to measure the statistical impact of Peer Review on student outcomes. Rather, it supports a conceptual and practice–oriented model of the organisational functions required to

make Peer Review work as an evidence–informed process. The data were used to reconstruct the sequence of activities involved in Peer Review, identify recurring organisational problems, understand the responsibilities that are often implicit, and define the conditions under which feedback can be transformed into organisational learning.

Methodology

The research design integrates documentary analysis, literature review and qualitative interviews. Documentary analysis was used to identify the formal architecture of Peer Review and quality assurance in VET: phases, expected outputs, decision points, evidence requirements and links with improvement planning. The literature review provided the theoretical categories needed to interpret Peer Review as a social and organisational process, rather than as a purely technical procedure.

The interviews were analysed through an approach combining task analysis, job analysis, behaviour analysis and the critical incident technique. Task analysis made it possible to reconstruct the activities that occur before, during and after Peer Review. Job analysis helped identify the functions and responsibilities required to perform these activities. Behaviour analysis focused on the practices that enable or obstruct participation, evidence use and feedback. The critical incident technique was used to recognise situations in which the process succeeds or fails: for example, when evidence remains fragmented, when responsibilities overlap, when external recommendations are not translated into action, or when professional dialogue becomes a trigger for improvement.

The objective was not to design a standard organisational chart. VET centres differ in size, resources, maturity of quality systems and internal professional cultures. For this reason, the analysis distinguishes between organisational positions and internal roles. A role is understood as a function that must be performed, even when it is not associated with a formal job title. In small centres, one person may perform several roles; in larger organisations, the same role may be distributed among different people. This flexible approach allows the model to identify key functions without imposing a rigid structure.

Results and discussion

The first result is the reconstruction of Peer Review as a four–phase process: design, self–evaluation, peer visit and redesign. The design phase defines objectives, quality areas, timing, resources, responsibilities and strategies for involving internal actors. The self–evaluation phase organises evidence collection, analysis of available data, consultation with stakeholders and preparation of the self–report. The peer visit introduces an external professional perspective through observation, dialogue and feedback. The redesign phase translates findings, recommendations and learning into an improvement plan with objectives, actions, responsibilities, timelines and monitoring indicators. This sequence shows that Peer Review does not end with the external feedback report. Its value depends on the organisation's capacity to rework what emerges and start a new cycle of improvement.

The second result is the identification of six internal roles: Director, Quality Manager, Facilitator, Research Expert, Improvement Designer and Specific Internal Actors. The Director provides strategic legitimacy, allocates resources and connects Peer Review to the organisation's quality system. The Quality Manager ensures methodological coherence, documentation, procedures and alignment with the internal quality management system. The Facilitator supports dialogue, participation, trust and the quality of internal interactions. The Research Expert contributes to the definition of evaluation questions, data collection, interpretation of evidence and reliability of the self–evaluation process. The Improvement Designer translates evaluation findings into feasible, monitorable and resource–sensitive actions. Specific Internal Actors, such as teachers, tutors, administrative staff and other operational figures, provide situated knowledge of processes and participate in both evidence collection and implementation of improvements.

The framework shows that organisational learning emerges from the interaction among these roles. The Director can provide mandate and resources, but without a Quality Manager the process may lose coherence. The Quality Manager can ensure procedures and documentation, but without a Facilitator the discussion may remain formal or defensive. The Research Expert can improve the quality of evidence, but without the contribution of Specific Internal Actors the data remain partial and detached from practice. The Improvement Designer can transform recommendations into actions, but without distributed involvement the plan may remain unimplemented. Peer Review therefore depends less on isolated individuals than on a system of distributed, interdependent responsibilities.

This finding is relevant to the Session 6 theme because it reframes data literacy and data use as organisational capacities, not only as individual skills. In many quality assurance processes, data are collected but not used to change practices. The proposed model suggests that evidence–informed improvement requires internal roles capable of connecting data, professional interpretation, decision–

making and follow–up. Data use is not a linear transfer from indicators to decisions. It is a situated process mediated by roles, professional cultures, power relations, available resources and organisational routines.
The model also highlights a change in the posture of the contemporary evaluator. The evaluator is not only a technical expert in measurement or a guarantor of compliance. In evidence–informed Peer Review, the evaluator acts as an architect of dialogic contexts. This means ensuring methodological rigour, correct use of evidence and traceability of decisions, while also enabling participation, feedback, negotiation of meaning and translation of recommendations into practical actions. Evaluation is therefore understood as a social process informed by data: evidence does not speak by itself; it must be constructed, interpreted and discussed within organisations.

Conclusions

The contribution proposes a framework of internal roles that can help VET centres use Peer Review as a driver of organisational learning and quality improvement. Its main contribution is to make visible the internal functions that are often implicit in Peer Review processes. By distinguishing roles from formal positions, the framework can be adapted to different organisational contexts while preserving the key responsibilities needed for evidence–informed improvement.
The model has implications for policy, practice and research. For policy, it suggests that quality assurance frameworks should not only define procedures and indicators, but also support organisations in developing the internal capacities required to use evidence. For practice, it offers VET centres a tool to prepare Peer Review, distribute responsibilities, avoid gaps or overlaps, and ensure continuity between evaluation and improvement planning. For research, it invites further empirical investigation into how internal roles, professional cultures and evidence use interact in different educational and training settings.
The paper does not present a counterfactual impact evaluation, and it does not claim that Peer Review automatically produces improvement. On the contrary, it argues that the effect of Peer Review depends on the organisational conditions under which it is implemented. When evidence is collected, interpreted and acted upon through distributed roles and professional dialogue, Peer Review can move beyond compliance and become a meaningful practice of organisational learning.

Keywords: Evidence–informed practice; Peer Review; Organisational learning; Quality assurance

What Data, for What, and Why? School and Teacher Data Use Beyond National Assessments in Italian Schools

Giulia Montefiore

Introduction

Expectations for schools to use data from multiple sources to improve the education they provide have been increasing globally (Verger & Skedsmo, 2021) and what counts as data in education has expanded beyond standardised national large–scale assessments. The term School Autonomy With Accountability (SAWA) indicates those educational policies that have become globally endemic, influenced by New Public Management and by the pervasiveness of technology and datafication in education (Verger et al., 2019). Centered on accountability and school autonomy to govern education systems, these policies generally make use of standardised national large–scale assessments (NLSA) to measure learning. Crucial is the role of the use of data, which has evolved from being merely focused on accountability, to being increasingly aimed at activating school improvement processes with a progressively expanding view of what data is and means.
In Italy, the developments of school autonomy since 1997, the National Evaluation System (SNV) and Law n.107 of 2015 configure the Italian education regime as close to SAWA, with particular attention to the continuous improvement of schools. Within the Italian school autonomy and accountability system, schools are expected to analyse and reflect on data provided by central administration as well as data produced locally to support pedagogical and organisational improvement. In fact, the SNV assigns centrality to schools' didactic and organizational improvement, and to schools' very own analysis and reflection about the data that they both receive from the central administration and produce locally.
The INVALSI Institute collects information on the use of data – especially relating to the results of national census–based learning assessments, the INVALSI tests – by annually administering a questionnaire to school principals and teachers of a representative sample of schools. However, it struggles to emerge how

much, how, why and for what purposes schools make use of INVALSI data. In 2016, Pastori and Pagani reported growing trust but difficulty in using national assessment data. Several years later, with a fully operational national school evaluation system, Italian schools appear to interact with data from multiple sources at varying levels of intensity. It is therefore necessary to analyze what is available and to research which elements favor the use of data in Italian schools, including and beyond data from INVALSI results.

This research takes the role that externally and internally provided data is expected to play, particularly in the absence of high stakes, and in connection to national tests and data use influence on pedagogical and assessment strategies, and investigates Italian SAWA policies *enactment*, that is in which circumstances and for whom these policies produce particular effects (Ball et al., 2012), especially in relation to the use of data in schools. This study investigates both the organizational level and the teacher individual level.

More specifically, the objective of the research is to understand how processes of data use take place in schools. It analyses how school organizational and political context, individual factors and different accountabilities mediate data use, and whether and how data influences organizational, pedagogical, and assessment strategies in schools. Data is not only that coming from national tests, but also that of the self–evaluation report and data internally produced by schools. Consequently, this research explores how schools respond to policy expectations, as to methods and outcomes, regarding data use. This study therefore investigates the extent to which data–use processes and routines are present in Italian schools, what actors consider to be data, and what attitudes they hold towards its role in organisational and pedagogical improvement. Attention is given to collaboration, collective reflection, and shared spaces and time within schools to understand how data are interpreted by school actors and how they contribute to decision–making processes. This study also analyses teacher's individual perceptions and reactions to expectations to use data from, specifically, the INVALSI tests to improve their own work as teachers, and how teachers use this data in their profession.

The theoretical perspective on SAWA and data use is informed by an understanding of policy instruments as not neutral (Lascoumes & Le Galès, 2007) and generating constituencies expanding instruments' realm of application (Simons & Voß, 2018). The study adopts an understanding of policy as enacted, interpreted and reconfigured by local actors rather than neutrally implemented (Ball et al., 2012). It draws on conceptualisations of data use in education as practices (Coburn and Turner, 2011) and as data not objectively guiding decisions (Spillane, 2012). It engages with the notion of data–informed rather than strictly data–based decision–making (Schildkamp et al., 2019), and considers the potential role of data use in supporting equity–oriented practices (Datnow & Park, 2018).

METHODOLOGY AND DATA USED

Methodologically, this analysis of enactment and data use employs a qualitative methodology. It draws on interviews (n=88) and focus groups (n=13) with principals, leadership teams, and teachers, and school document analysis in ten comprehensive primary and lower–secondary schools in Rome, of different socio–economic contexts.

Schools were selected from the smallest administrative areas to which each school would belong, the Urban Planning Zone (Zona Urbanistica), using the Social Hardship Index (Indice di disagio sociale – IDS), developed by the city of Rome based on 2011 census data. The 2011 IDS data was cross verified with schools' 2021 self–evaluation reports where schools narratively describe their socio–economic context to ensure current relevance. Schools were mapped by percentage of lower–secondary students scoring in the two lowest bands (out of five) in the 2021/22 Italian National Large–Scale Assessments, the INVALSI tests, using the Transparent School (Scuola in Chiaro) Ministry of Education online platform. The objective was to select pairs of schools belonging to the same Local Education Market (LEM) but differing in national tests results, to allow comparison within the same LEM but also between schools of similar socio–economic background or INVALSI results in different areas. Based on this process, participation of 10 schools was secured.

The semi–structured interview protocol was developed both based on literature conceptualisations of data use (Coburn and Turner, 2011) and on pre–existent empirical work. The protocol captures: respondent's individual characteristics; school context; and school organisational characteristics, both general and directly in relation to elements deemed relevant in explaining data use practices in schools. The interview protocol focuses on different data sources, namely INVALSI national large–scale assessment data, self–evaluation data, grade–based school–wide tests, and asks about a few other sources of data. For each data source the same elements of data use models are investigated. The focus groups aim at exploring teachers' individual experiences of, this time, only INVALSI tests results, and were designed to explore their experiences in reference to expectations to use this data, and their approaches to and practices of data use.

For the analysis, the research employs a mix of inductive and deductive coding strategy, based on the existing literature of data use while also staying open to the emergence of new codes during data analysis.

RESULTS & DISCUSSION

Preliminary findings suggest a limited and heterogeneous presence of structured data–use practices. Here, organisational constraints and uneven levels of data literacy appear to constitute important barriers to sustained engagement with data. The production and engagement with self–evaluation data is also mediated by organizational factors or organizational choices that either enable or hinder self–reflection practices.

Often, data provide evidence for already known issues and are experienced as legitimising the necessary action to address them. For example, INVALSI and grade–based school–wide tests may provide what is perceived as 'objective' evidence, to address sustained and previously known issues with specific teachers in terms of their pupils' academic achievement.

Less often, data are used to inform decision–making that fully exploits school autonomy to address equity issues. This is particularly evident in decisions concerning class composition and organizational arrangements of lower–secondary education. In some schools, principals, based on INVALSI results, changed the usual class formation practice based on the students' second foreign language selection. Instead, they created more socially and academic heterogenous classrooms, where students would be subsequently regrouped for the second foreign language instruction. Principals used data to inform decisions, available to them through school autonomy, to counteract processes of academic and social stratification.

Still, while many schools express interest in using data to better understand student needs and inequalities, systematic processes supporting equity–oriented data use remain unevenly embedded in school practices. This may be due, amongst other factors, to perception of viability, leadership approaches and lack of data literacy.

In regard to individual teachers' experience of expectations to use data from INVALSI tests and their use of data, experiences differ between primary and lower–secondary teachers. The relationships of these two group of teachers with data may vary also, albeit not only, due to the granularity of data they can access after testing and the possibility to review exactly what their students have been assessed on, in terms of test content. When data use occurs, in the case of primary school teachers, it may not necessarily happen in the systematised way policy would expect, but more via direct engagement with test content from the official paper–based tests and by reviewing teachers' own pupils' results. Interestingly, attitude towards data use may change over the course of the interviews and focus groups, as both primary and lower–secondary teacher respondents reflect on their own experiences more deeply than previously done, learn about their own colleagues' experiences, and discuss the topic with other teachers. During the research it emerged that often, the organisational elements of dedicated time and space, and organisational routines to talk about and reflect on data had been scarce or not present, in the respondents' school experience.

CONCLUSION

This study contributes empirical evidence on data use practices in a Southern European low–stakes accountability context still relatively underexplored in the literature. It advances understanding of how organisational conditions, data literacy, and collaborative practices shape data–informed decision–making and the potential use of data for equity–oriented purposes. The research explores determinants and possibilities in regard to teachers' individual data perceptions and data use practices.

This research results and respondents' feedback on the usefulness of interviews and focus groups specifically on the topic of data use, opens the possibility for further research in the future aimed at enabling schools and school actors to engage in the data use they wish they could undertake, in reference to INVALSI data, grade–based school–wide test data, and data emerging from school self–evaluation.

Keywords: Data use; Enactment; Data literacy; Sensemaking

Motivated but Challenged: Measuring Attitudes Toward Data Science Among Graduate Students in Educational Research Methodology

Ioulia Televantou – Maria Meletiou – Georgia Solomonidou – Daniel Frischemeier

Introduction

Introductory statistics and research methods courses in the education sciences have traditionally relied on relatively small, tidy, and structured datasets, often selected to help students learn statistical concepts through meaningful educational and social contexts (Engel, 2017). However, the data now available to educational researchers and professionals increasingly differ from these pedagogical datasets in scale, complexity, format, and origin. Alongside survey and observational data, educational systems now generate data through digital platforms, web–based learning environments, learning management systems, digitized large–scale assessments, and administrative infrastructures (McFarland et al., 2021; Romero & Ventura, 2020). In response to these developments, Educational Data Science (EDS; McFarland et al., 2021; Peña–Ayala, 2023) and Learning Analytics (LA; Long & Siemens, 2011) have emerged as fields concerned with making sense of rich, complex, and often unstructured educational data sources.

In this context, the present paper addresses two interrelated gaps: a curriculum gap and a measurement gap. First, it identifies educational data science as an additional methodological strand within educational research, extending traditional qualitative, quantitative, and mixed–methods approaches to include data–intensive, computational, and analytics–based forms of inquiry (McFarland et al., 2021; Piety et al., 2014). Second, before this curriculum gap can be meaningfully addressed, a prior measurement gap must be resolved. Evidence from statistics education suggests that learners' attitudes are important antecedents of engagement, persistence, learning, and literacy development (Ramirez et al., 2012). By extension, attitudes toward data science may constitute an important foundation for educational data science literacy (Kerby–Helm et al., 2022). Nevertheless, research on data science attitudes remains limited, particularly among education and social science students enrolled in Research Methodology courses and in non–English–speaking contexts.

Against this background, the present study adopts Expectancy–Value Theory (Eccles et al., 1983; Eccles & Wigfield, 2002, 2020) as a framework for conceptualizing attitudes toward data science and contributes to this emerging field by examining the psychometric properties of a multidimensional measure of data science attitudes, namely the Student Survey of Motivational Attitudes toward Data Science (S–SOMADS; Kerby–Helm et al., 2022; Kerby–Helm et al., 2024; Whitaker et al., 2024), among Greek–speaking Educational Research Methodology students.

The central argument is that closing the measurement gap is necessary before the curriculum gap can be systematically addressed. Without valid instruments, it is difficult to examine how students perceive data science, whether they feel confident and interested in engaging with it, or how they respond to curricular attempts to introduce educational data science. This is particularly important in educational research methodology contexts, where students may encounter data science as an emerging extension of research training rather than as a stand–alone technical field.

The research questions guiding the study are the following:

(1) What measurement structure provides the most theoretically coherent and psychometrically acceptable representation of data science attitudes, as measured by the adapted Greek S–SOMADS, among Educational Research Methodology students?

(2) To what extent are data science attitudes associated with students' demographic characteristics, academic background, and prior quantitative or technological preparation?

Methodology

Instructional Context and Procedure

Ethical approval for the main validation study was obtained from the Cyprus National Bioethics Committee (ΕΕΒΚ ΕΠ 2025.01.367). As part of the instructional context, students attended a three–hour seminar titled *Data Science and Applications in Educational Research*. The seminar introduced survey data, large–scale assessment data, process and log data, combining conceptual explanation with interactive activities and reflection questions, and encouraging students to consider how educational data science could inform their own research and practice. It also addressed ethical and practical challenges, including data access, data cleaning, privacy, algorithmic bias, and the need for appropriate statistical and computational skills. Following the seminar, students were invited to complete an online questionnaire administered through Microsoft Forms. A response window of 10 days was provided. Importantly, the seminar is treated in the

present study as the instructional context in which attitudes toward data science were assessed, rather than as an intervention whose causal effect is evaluated. It also served a methodological function: by providing students with an introductory understanding of data science before they completed the adapted S–SOMADS, it helped ensure that their responses referred to data science specifically, rather than to research methods, statistics, ICT, or educational technology more generally. The final version of the instrument was obtained through personal communication with the S–SOMADS development team. For the purposes of the present study, the S–SOMADS was adapted into Greek through a multi–stage translation and adaptation procedure.

Statistical Analysis

Analyses were conducted using the lavaan package (Rosseel, 2012). Following a staged validation approach (AERA et al., 2014; Messick, 1995), we examined descriptive statistics, missing–data patterns, and the measurement structure of the adapted Greek S–SOMADS using confirmatory factor analysis (Kerby–Helm et al., 2022, 2024). Missing item responses were handled with full information maximum likelihood (Enders, 2010). Model refinements were retained only when psychometrically improved and theoretically interpretable (Brown, 2015; Kline, 2023). Reliability was assessed using Cronbach's alpha and McDonald's omega (McDonald, 1999). Finally, associations with students' prior quantitative, technological, and academic preparation were examined as evidence based on relations with external variables (AERA et al., 2014).

Results

Analytic Sample and Participant Characteristics

Participants ranged in age from 22 to 59 years (M = 34.85, SD = 10.14). The majority were female (85.6%) and resided in Greece (89.5%), with the remaining participants residing in Cyprus (10.5%). Most participants reported previous coursework in Information Technology (91.3%), while approximately two–thirds had previous coursework in Statistics (62.8%). Most participants were enrolled in the Educational Research Methodology course (75.4%), whereas 24.6% attended the Quantitative Methods in Educational Sciences course. Regarding prior academic achievement, 76.3% reported a high school graduation grade between 16 and 20, while 23.7% reported a grade below 16. Most participants had a non–STEM academic background (85.9%), whereas 14.1% had a STEM background.

Evaluation of the Original S–SOMADS Subscale Models

Table 1. *Construct Validity and Reliability Evidence (Final Scales)*

Construct	χ^2(df)	CFI	TLI	RMSEA [90% CI]	SRMR	α	ω
Technical & Critical Expectancy	63.06 (19)	.934	.903	.073 [.057, .090]	.046	.850	.876
Competence & Identity Attainment	54.63 (8)	.934	.877	.115 [.090, .141]	.052	.788	.803
Investment Cost	10.28 (2)	.928	.783	.099 [.047, .160]	.036	.806	.813
Difficulty	18.71 (9)	.975	.958	.050 [.021, .078]	.038	.847	.852
Interest/Enjoyment	60.80 (32)	.966	.952	.046 [.033, .059]	.048	.686	.764
Intrinsic Goals	5.22 (2)	.984	.953	.063 [.014, .112]	.031	.828	.824
Utility	17.65 (9)	.973	.956	.048 [.014, .078]	.032	.910	.930

Note. Robust MLR estimation with FIML for missing data. CFI = Comparative Fit Index; TLI = Tucker–Lewis Index; RMSEA = Root Mean Square Error of Approximation; SRMR = Standardized Root Mean Square Residual. α = Cronbach's alpha; ω = McDonald's omega total. Omega total is reported because all retained scales were modelled as unidimensional factors for reliability estimation. Values ≥ .70 are generally considered acceptable for research purposes. Data derived from the final retained measurement models.

As a first step, the measurement model for each original S–SOMADS construct was evaluated separately using confirmatory factor analysis. This stage examined whether the item sets proposed in the original instrument could be adequately represented as unidimensional constructs in the present sample of Educational Research Methodology students. The results indicated that the original unidimensional structure was not equally supported across all constructs.

Subsequent analyses therefore examined whether theoretically interpretable refinements could improve the measurement structure of each construct. Expectancy was better represented by two related dimensions, reflecting confidence in performing data science tasks (Technical Expectancy) and confidence in critically evaluating data–based evidence (Critical Expectancy). This distinction is substantively important because it suggests that students differentiated between "doing" data science tasks and "judging"

the quality, validity, and ethical implications of data–based claims. Attainment was also better represented by two related dimensions, distinguishing competence–based attainment value (Competence Attainment) from identity–based attainment value (Identity Attainment). Evidence for constructing validity and reliability was generally supportive of the retained measurement model (Table 1).

Descriptive Statistics & Associations with Participants Characteristics

Students reported the highest mean for Competence Attainment, indicating strong links between data science success, satisfaction, and pride. Scores were also relatively high for Identity Attainment, Interest, Intrinsic Goals, Investment Cost, and Critical Expectancy, suggesting that students valued data science and saw it as meaningful. However, lower scores for Utility and Technical Expectancy, together with above–midpoint Difficulty, indicate weaker perceived career value, lower technical confidence, and a view of data science as challenging. Overall, students appeared motivated and value–oriented, but also aware of the demands of learning data science.

External validity evidence was examined by analyzing associations between participant characteristics and the retained data science attitude dimensions (see Table 2). Findings significant at the conventional $\alpha = .05$ level but not after Bonferroni correction were interpreted as tentative. No statistically significant differences were observed by gender or country of residence after correction for multiple testing. Similarly, attitudes toward data science were largely comparable across the two course contexts. In contrast, indicators of prior quantitative and academic preparation showed more consistent associations with competence–related dimensions. Participants with higher high–school mathematics achievement, previous statistics coursework, and with STEM degree backgrounds reported higher Technical Expectancy and lower perceived Difficulty, with associations remaining statistically significant even after Bonferroni correction.

Multiple regression models were estimated for Technical Expectancy and Difficulty to examine whether associations with prior quantitative preparation remained after controlling for participant characteristics. Both models were statistically significant. High–school mathematics achievement was the only significant and unique predictor of Technical Expectancy, with higher achievement associated with greater technical confidence. For Difficulty, STEM degree background and higher high–school mathematics achievement predicted lower perceived difficulty, with STEM background showing the stronger association.

Discussion

This study addressed a measurement gap in educational research methodology curriculum by adapting and examining the psychometric properties of the Greek version of the Student Survey of Motivational Attitudes toward Data Science (S–SOMADS; Kerby–Helm et al., 2022) among postgraduate education students enrolled in Educational Research Methodology courses. The Greek S–SOMADS does not merely translate an existing instrument; it provides insights into how education students organize their attitudes toward data science when data science is introduced as one component of broader research methodology preparation, rather than as the central focus of a dedicated data science course or program of study. This makes the instrument useful not only for measurement, but also for curriculum design, because it identifies where students feel capable, where they perceive difficulty, and how they value data science as part of their professional and academic development.

From a practical perspective, these findings support the use of the adapted S–SOMADS as an accessible formative tool for instructors and program designers. From a substantive perspective, the distinction between valuing data science and feeling technically capable of engaging with it is pedagogically important. Curriculum reform should not assume that positive motivational value is sufficient for successful engagement with Educational Data Science. Students may value data science and consider it relevant to educational research, while still feeling uncertain about their ability to perform technical procedures, work with data, or interpret statistical results. In this sense, the findings suggest that the main challenge for many education students may not be convincing them that data science matters but helping them develop confidence and reduce perceived difficulty.

Keywords: Educational data literacy; Educational research methodology; S–SOMADS; Expectancy–Value Theory

SESSION 10. UNDERSTANDING ONE–YEAR ACHIEVEMENT GROWTH IN MATHEMATICS: EVIDENCE FROM TIMSS LONGITUDINAL AND ITS RELEVANCE FOR ITALY

ORGANIZER: IEA - AARHUS UNIVERSITY
COORDINATORS: NURULLAH ERYILMAZ – ROLF STRIETHOLT
6TH NOVEMBER: 11.00 A.M. – 1.00 P.M. {AULA A1.1 – RESEARCH 20}

Changes in response time and mathematical performance in TIMSS–L

Jose G. Clavel – Matthias von Davier – Mao Thianzheng

Introduction

Assessing the extent of students' academic growth over a year is crucial to understanding educational outcomes. However, measuring change implies that the same students are assessed repeatedly over time. Even though students are tested again after a break of almost a full year of schooling, putting the same students into the same situation again may have an effect on test–taking behaviors. This is particularly likely in a low–stakes International Large Scale Assessment (ILSA), where students do not face any consequences related to their test performance. Students who exert low effort as a consequence may answer items too quickly, engage in rapid guessing, or idling (Bulut et al., 2026; Shute et al., 2016; Wise & Kong, 2005). Therefore, it is important to consider how changes in response behavior are associated with changes in observed academic achievement. In digital assessments like TIMSS 2023 and Longitudinal, these changes are reflected in achievement growth and response behaviors, such as item response time (Anghel, 2023).

Response behavior and response time provide critical information for interpreting student achievement, as they reflect the processes underlying observed item responses (Pohl, Ulitzsch & von Davier, 2021). These indicators enable researchers to better disentangle differences in disengagement and response behavior from achievement estimates (Leng et al., 2025). In addition, process data offer fine–grained information for examining response patterns in complex problem–solving tasks, such as TIMSS Problem Solving and Inquiry (PSI) items (Tan & Bulut, 2025; Bulut, 2025), and for investigating position effects in item responses (Liu et al., 2024).

Response time is associated with various item characteristics, including difficulty, length, position, complexity and item type (Halkitis et al., 1996; Zenisky & Baldwin, 2006). The association between response time and item characteristics was found to be moderated by students' ability (Goldhammer et al., 2014; Kroehne et al., 2020; Michaelides et al., 2020). In addition, students' response behaviors may vary across different backgrounds and environmental factors, suggesting that contextual variables also shape how students engage with assessment items (Jensen et al., 2018; Soland & Kuhfeld, 2019). Within the TIMSS framework, process data have also supported a range of research applications, including improving the interpretation of achievement gaps (Leng et al., 2024), refining scale estimation (Leng & von Davier, 2026), examining response behavior in complex problem–solving tasks (Tan & Bulut, 2025; Bulut, 2023) and investigating position effects in item responses (Liu et al., 2024). Instead of focusing on analysis within a single test administration, our study aims to examine how changes in response behavior are associated with estimated changes in student math achievement (growth).

While prior research has established associations among response time, item characteristics and student ability, these studies are based solely on data from a single assessment administration. Consequently, limited evidence exists on how within–student changes in response behavior relate to observed achievement growth over time, particularly in large–scale low–stakes assessments. This raises an important question: do observed achievement gains reflect substantive learning, changes in test–taking behavior, or a combination of both? To analyze such changes over time, the data should be collected from an assessment that is administered under comparable testing conditions and to comparable samples.

In 2024, TIMSS introduced its first longitudinal study by extending the TIMSS 2023 main study. The longitudinal component is designed to provide high–quality data for tracking students' growth over a one–year interval. In addition, it collects student–, teacher– and school–level background information to offer insights into the learning context and environmental factors associated with student development (von Davier et al., 2025). The same cohort of students is re–assessed to ensure comparability over time. The assessment booklets used during the reassessment are the same as those used in TIMSS 2023, enabling

TIMSS–L to produce comparable achievement estimates using the same item parameters estimated from TIMSS 2023. To mitigate potential item exposure effects, a special rotation is used that ensures students receive different booklets in the first and second assessment (Fishbein et al., 2025), thereby avoiding repeated exposure to the same items. This procedure is only applicable to the performance items: the background questionnaires used in the reassessment are nearly same as those used in the TIMSS 2023 main study, thus allowing for the examination of within–student changes over time and helping reduce the influence of between–student differences when interpreting observed growth.
In addition, TIMSS has begun releasing process data from the 2019 and 2023 cycles, as well as from the 2023 longitudinal extension (Fishbein et al., 2025). These data constitute a key resource for examining students' response behaviors and their relationship to estimated achievement. The longitudinal design further enables linking process data across both the main study and the follow–up assessment for the same students, thereby supporting the analysis of within–student behavioral changes over time. This provides additional insight into how students' interaction patterns with assessment tasks may evolve over a one–year period within different educational contexts.

Keywords: Large Scale Assessment; Process data; Context Scale; Linear Regression Analysis; Cognitive Domain; Achievement Growth

Understanding One–Year Achievement Growth in Mathematics: Evidence from TIMSS Longitudinal and Its Relevance for Italy

Nurullah Eryilmaz – Rolf Strietholt

TIMSS Longitudinal provides a unique opportunity to investigate mathematics achievement growth over one year of schooling by reassessing the same students who participated in TIMSS 2023 one year later in 2024. As the first student–level longitudinal extension of TIMSS, the study enables researchers to move beyond cross–sectional comparisons of achievement levels and examine learning gains, inequalities in progress, and the conditions under which students develop mathematical competencies over time.
This symposium brings together five coordinated papers that leverage TIMSS Longitudinal data to examine one–year mathematics achievement growth from complementary perspectives. Together, the papers investigate how individual student characteristics, teacher qualifications, instructional quality, school climate, and assessment process data contribute to understanding students' learning trajectories. Particular attention is given to Italy's participation in the Grade 4–5 longitudinal design alongside eight other education systems, while comparative insights are also drawn from the Grade 8–9 design, in which three education systems participated. In doing so, the symposium contributes to the conference theme by demonstrating how international large–scale assessment data can inform research and policy discussions on educational trajectories, equity, educational effectiveness, and the interpretation of longitudinal learning outcomes.
Across the five papers, the symposium addresses two overarching questions: (1) Which student, teacher, instructional, and school factors are associated with stronger mathematics achievement growth? and (2) To what extent do assessment process indicators improve our understanding and interpretation of observed learning gains? Methodologically, the papers employ multilevel longitudinal models and process–data analyses that account for the nested structure of schooling while exploiting the unique longitudinal design of TIMSS. Conceptually, the symposium integrates perspectives from cumulative advantage theory, teacher quality research, instructional effectiveness, school climate research, and educational measurement to provide a comprehensive understanding of mathematics achievement development and educational inequality.
The first paper investigates the role of student characteristics in shaping mathematics achievement growth, with particular attention to cumulative advantage (Matthew effects) and differences associated with gender, socioeconomic status, and migration background. The second paper examines whether teacher qualifications and professional preparation predict students' learning growth, contributing to ongoing debates regarding the extent to which formal teacher characteristics translate into improved student outcomes. The third paper focuses on instructional quality, operationalized through classroom management, cognitive activation, and supportive climate, and evaluates its association with mathematics achievement growth. The fourth paper examines the school as an organization, analysing how school

climate—including safety, bullying, and academic emphasis—relates to achievement growth and whether positive school environments mitigate or reinforce socioeconomic inequalities in learning trajectories. The fifth paper complements these substantive analyses by investigating changes in students' response times and test–taking behaviour, examining how process data contribute to the interpretation of observed mathematics achievement growth in low–stakes international assessments.
TIMSS Longitudinal includes achievement data together with student, home, teacher, school, and national context information collected in 2023 and 2024, making it particularly well suited for integrated analyses across micro–, meso–, and macro–level contexts. By combining substantive analyses of learning processes with methodological advances in longitudinal modelling and process–data analysis, the symposium illustrates the broad analytical potential of TIMSS Longitudinal for advancing research on educational effectiveness, educational equity, and student learning across countries.
By bringing together researchers from educational measurement, sociology of education, economics of education, psychology, statistics, and comparative education, the symposium offers a multidisciplinary discussion of how longitudinal international assessment data can advance understanding of mathematics achievement growth, educational effectiveness, and equity in Italy and beyond.

Keywords: Longitudinal assessment; Mathematics achievement growth; Comparative education

Individual Student Characteristics and Mathematics Achievement Gains: Evidence from the TIMSS Longitudinal Study

Nurullah Eryilmaz – Rolf Strietholt – Trude Nilsen – Stefan Johansson

Longitudinal research in educational inequality aims to understand how earlier achievement and social background shape subsequent learning trajectories. While prior achievement is a strong predictor of later performance, theoretical perspectives diverge on whether it leads to cumulative advantage—where high–achieving students experience greater gains—or compensatory patterns, where lower–achieving students show relatively stronger growth. Empirical evidence from large–scale studies has often challenged strong cumulative advantage assumptions, instead pointing to modest or absent Matthew effects in reading and mathematics. These patterns are frequently attributed to regression–to–the–mean dynamics, instructional differentiation, or ceiling effects (Baumert et al., 2012; Nennstiel, 2023). Beyond prior achievement, social background characteristics like socioeconomic status (SES) play a key role in shaping outcomes. SES captures inequalities in access to material resources, cultural capital, and learning support. While SES differences in achievement levels are well–documented, longitudinal research suggests that growth gaps are often stable rather than widening over time once prior achievement is taken into account. Evidence from Germany, for example, indicates persistent SES differences in mathematics and language levels but little support for cumulative SES–related divergence in trajectories (Nennstiel, 2023). Examining SES effects net of prior achievement allows for a precise assessment of whether students with similar starting points diverge. Gender differences in mathematics achievement represent another longstanding topic. Cross–national evidence shows that gender gaps vary considerably across countries and grade levels. Longitudinal analyses suggest these differences tend to remain relatively stable over time, with only modest changes during secondary education. Recent evidence indicates that gender differences in mathematics development are generally small once prior achievement is controlled for, though persistent patterns may still be observed in specific systems or cohorts (Uunk & Blossfeld, 2025). Controlling for prior achievement provides a framework for assessing whether gender is associated with differential achievement outcomes beyond earlier performance. This study addresses these issues by drawing on the complete longitudinal dataset from the TIMSS Longitudinal Study. The analysis encompasses a primary cohort following students from Grade 4 to Grade 5 across nine countries, and a lower secondary cohort following students from Grade 8 to Grade 9 across three countries. Focusing on mathematics, the study employs two–level regression models with students nested within schools to examine achievement at the second measurement occasion while explicitly controlling for prior achievement. By investigating prior achievement, gender, and SES, this study provides a baseline individual–level model of achievement growth.

Keywords: Longitudinal assessment; Mathematics achievement growth; Comparative education

School Climate: Impact of Safety, Relationships, Resources, and Academic Emphasis on Achievement Gains and Equality: TIMSS Longitudinal across countries and grades

Rolf Strietholt – Trude Nilsen – Stefan Johansson – Nurullah Eryilmaz

This study investigates which aspects of school climate relate to progress in student learning outcomes and whether students have equitable access to a positive school climate. School climate is a multifaceted concept; systematic reviews identify four key dimensions: academic (promotion of learning), community (interpersonal relationships), safety (environmental security and absence of bullying), and institutional (organizational structure and resources) (Wang & Degol, 2016; Thapa et al., 2013). While these dimensions are critical to cognitive and affective outcomes (Cohen et al., 2009), it remains unclear which specific dimensions most significantly drive learning progress and educational equality (Berkowitz et al., 2017). Current research often focuses on single dimensions, and findings remain inconsistent due to varying conceptualizations and methodological limitations (Nilsen & Teig, 2022). Furthermore, most studies rely on cross–sectional designs prone to omitted variable bias and reversed causality, hindering robust causal inferences. Consequently, there is an urgent need for longitudinal research using representative samples and validated frameworks that allow for cross–national comparisons. This study addresses these gaps by utilizing the complete longitudinal dataset from the TIMSS Longitudinal project and adopting the framework by Wang and Degol (2016). The analysis encompasses all participating countries across two cohorts: a primary cohort (Grade 4 to 5) across nine countries, and a lower secondary cohort (Grade 8 to 9) across three of these countries. Using two–level regression analyses (students nested within classes), the study investigates the impact of the four school climate dimensions on mathematics achievement gain. Additionally, it examines the relationship between student socioeconomic status and school climate to determine if positive school climates are equitably distributed. Missing data will be handled using robust maximum likelihood (RML) estimation to ensure the integrity of the longitudinal findings. The data will be available in February 2026. Based on previous research synthesis (e.g. Nilsen and Teig, 2022; Wang and Degol, 2016), we hypothesize that absence of bullying (safety dimension) and academic emphasis (academic dimension), alongside student sense of belonging (community dimension), will demonstrate stronger impacts on cognitive outcomes than physical resources (institutional dimension). However, we expect results to vary across countries and grades. We further hypothesize that the equitable distribution of a positive school climate will vary according to the broader educational equity profiles of the participating countries (Berkowitz et al., 2017). These findings will offer evidence–based insights for school leaders and policymakers regarding the role of school climate in fostering both academic excellence and social equality.

Keywords: Longitudinal assessment; Mathematics achievement growth; Comparative education

Teacher Quality Indicators and Student Achievement Gains: Evidence from the TIMSS Longitudinal Study

Stefan Johansson – Rolf Strietholt – Nurullah Eryilmaz – Trude Nilsen

There is growing evidence that knowledge and skills acquired in compulsory school have lasting effects into adulthood (Gustafsson, 2016), and that the first years of schooling are particularly important for students' later educational trajectories. This raises the question of which early school determinants are most influential for student achievement. Among school resources, class size has been shown to affect achievement (Fredriksson et al., 2013), although effect sizes are modest. In contrast, several syntheses conclude that teaching and teachers are of greater importance (Hattie, 2009). Longitudinal studies consistently demonstrate substantial differences in student outcomes across teachers (e.g., Nye et al., 2004), with teacher–level variation accounting for approximately 10 percent of the variance in achievement. At the same time, evidence linking observed teacher characteristics such as education and experience to student outcomes has been mixed, partly due to methodological limitations and imprecise definitions of teacher qualifications (Blömeke & Olsen, 2019). Nevertheless, several studies report positive associations between teacher quality indicators and student achievement (Wayne & Youngs, 2003; Johansson et al., 2023). There is, moreover, a need for robust evidence, as much previous research relies on single–country studies or

cross–sectional data. Building on this literature, the present study examines variation in teacher effectiveness within schools and tests the hypothesis that teachers with relevant teacher education and longer experience have a greater impact on student learning than teachers without such qualifications. This study addresses these gaps by utilizing the complete longitudinal dataset from the TIMSS Longitudinal project. The analysis encompasses all participating countries across two cohorts: a primary cohort (Grade 4 to 5) across nine countries, and a lower secondary cohort (Grade 8 to 9) across three of these countries. Using two–level regression analyses (students nested within classes), the study investigates the impact of teacher quality indicators, such as teacher education and experience, on mathematics achievement gains. In addition, the study examines the relationship between student socioeconomic status and teacher quality to assess whether teacher quality is equitably distributed. We hypothesize that the distribution of teacher quality varies across countries as a function of the compensatory mechanisms of the school system. These findings provide evidence on the role of teacher quality in promoting student achievement and shed light on social selection effects.

Keywords: Longitudinal assessment; Mathematics achievement growth; Comparative education

The Importance of Teaching Quality for Mathematics Achievement Gains and Educational Equality: TIMSS Longitudinal data across all countries and grades

Trude Nilsen – Stefan Johansson – Rolf Strietholt – Nurullah Eryilmaz

This study investigates which aspects of teaching quality (TQ) are related to progress in student learning outcomes and whether students have equal access to high–quality instruction. Evidence identifies TQ as a fundamental driver of student learning and a critical link to educational equity (e.g. Fauth et al., 2021; Darling–Hammond, 2000). However, it remains unclear which specific TQ dimensions matter most for learning progress, as findings vary across countries and cohorts (Blömeke et al., 2019). These inconsistencies are often attributed to varying conceptualizations of TQ and methodological limitations (Klieme & Nilsen, 2022). While qualitative studies offer deep insights, they lack generalizability. Conversely, quantitative research frequently relies on cross–sectional designs prone to omitted variable bias and reversed causality, hindering robust causal inferences. Consequently, there is an urgent need for longitudinal research using representative samples and validated frameworks that allow for comparisons across countries and cohorts. This study addresses these gaps by utilizing the complete longitudinal dataset from the TIMSS Longitudinal project and adopting the Three Basic Dimensions framework (Klieme et al., 2009). This framework, comprising classroom management, supportive climate, and cognitive activation, provides a validated, standardized basis for comparison (Praetorius et al., 2018; Klieme & Nilsen, 2022). The analysis encompasses all participating countries across both available cohorts: a primary cohort (Grade 4 to 5) across nine countries, and a lower secondary cohort (Grade 8 to 9) across three of these countries. Using two–level regression analyses (students nested within classes), the study investigates the impact of the three TQ dimensions on mathematics achievement gain. Additionally, it examines the relationship between student socioeconomic status and TQ to determine if high–quality teaching is equitably distributed. Missing data will be handled using robust maximum likelihood (RML) estimation to ensure the integrity of the longitudinal findings. The data will be available in February 2026. Based on previous research (e.g. Praetorius et al., 2018; Klieme & Nilsen, 2022), we hypothesize that classroom management and cognitive activation will demonstrate stronger impacts on cognitive outcomes than supportive climate. We further hypothesize that the equitable distribution of high–quality teaching will vary according to the broader educational equity profiles of the participating countries. These findings will contribute to the global debate regarding the role of teaching quality in fostering both academic excellence and social equality**.**

Keywords: Longitudinal assessment; Mathematics achievement growth; Comparative education

SESSION 16. EDUCATIONAL INEQUALITIES, LEARNING OUTCOMES AND OPPORTUNITIES IN CHANGING CONTEXTS

ORGANIZERS: UNIVERSITY OF BERGAMO – UNIVERISITY OF MILAN "LA STATALE"
COORDINATORS: JURGENA MYFTIU – GIULIANA PARENTE
6TH NOVEMBER: 11.00 A.M. – 1.00 P.M. {AULA B0.1 – RESEARCH 21}

COVID–19, socioeconomic status, and upper–secondary track choice

Francesca Carimando – Loris Vergolini

Introduction

The COVID–19 pandemic produced one of the most substantial disruptions to education in recent history, interrupting schooling for students across the globe. In Italy, all schools were closed from March 2020 to the end of that academic year, and the two years that followed (2020/21 and 2021/22) were affected by locally differentiated restrictions and periods of remote or hybrid teaching. Although much research on the pandemic has focused on learning losses, considerably less attention has been paid to how the crisis may have shaped educational choices at decision points (Aalto et al., 2023; van de Werfhorst et al., 2023).

In the Italian educational system, students make a track choice in grade 8 (at age 13–14), choosing among three pathways: academic (*liceo*), technical (*istituto tecnico*), and vocational (*istituto professionale*). In the pandemic context, the loss of face–to–face interactions with teachers and counsellors, together with broader uncertainty about the future and the labour market, may have shifted how students and families assessed the value of different educational pathways. Pupils coming from a family with a low socioeconomic status may have felt these pressures more than their higher socioeconomic peers, facing additional challenges such unequal access to digital or home resources, among other factors.

Literature review

Socioeconomic background is a central dimension of educational inequality. Children coming from a high social class often benefit from more advantaged home learning environments, greater parental involvement, and more enriching out–of–school experiences compared to their less advantaged peers (Breen et al., 2009; Breen et al., 2010; Barone and Ruggera, 2018). Building on this, schools have proven to either reduce or reproduce these disparities: while some studies support the idea that schooling has a compensatory role, recent evidence suggests that this equalising effect may be limited or context–dependent (Passaretta & Skopek, 2021; von Hippel et al., 2018).

The COVID–19 pandemic may therefore have affected SES inequalities in two different ways. On the one hand, school closures could have reduced the advantage normally provided by higher–quality school environments, potentially producing larger learning losses among students who benefited most from schooling before the pandemic. On the other hand, reduced schooling increased the importance of the home learning environment, where high–SES families were generally better positioned to provide educational support, digital resources, and supervision during remote learning (OECD, 2020; Francis & Weller, 2022). For this reason, the pandemic may have widened SES gaps, especially where family resources became more central to maintaining learning.

Theoretical framework

From a rational–action perspective, educational decision–making (in this case, the choice of upper–secondary track) can be understood as a balance between perceived costs, expected returns, and risk preferences. External shocks like the COVID–19 pandemic alter perceptions of future returns and modify risk preferences, reshaping how families assess the value of academic versus more practically oriented pathways. At this regard, Ayarza (2023) conceptualises choices under uncertainty as a function of expectations, information, and non–pecuniary motives such as self–efficacy and sense of belonging. COVID–19 can thus be understood as a shock that reconfigured how students and families evaluated different educational trajectories. International evidence supports this hypothesis. In Sweden, the onset of the pandemic corresponded with a significant shift toward upper–secondary programmes associated with labour–market stability, notably in healthcare and technology (Aalto et al., 2023). Similarly Breidenbach and colleagues (2023) report that in Germany transitions to *Gymnasium* increased among students from high–income neighbourhoods but did not in low–income ones, suggesting an amplification of pre–existing disparities.

Transposed to the Italian context, it is possible that COVID–19 reinforced utilitarian educational choices, particularly among socioeconomically disadvantaged families. For households where fast entry into paid work may outweigh long–term academic aspirations, the uncertainty brought by the pandemic may have made technical and vocational tracks appear more rational. Building on this, the present study asks: *did the COVID–19 pandemic alter upper–secondary track choices in Italy, and did these changes differ between students with a different socioeconomic background?*

Data

The study draws on population data from the Italian National Institute for the Evaluation of the Education System (INVALSI–SNV), which conducts compulsory, low–stakes census tests in reading and mathematics each Spring for Grades 2, 5, 8, 10, and 13. For this study, INVALSI–SNV data have been longitudinally linked via a unique student identifier, allowing to follow pupils from lower secondary (Grade 8) into upper secondary (Grade 10) school. The analytical sample includes 443,515 observations across 5,548 schools.

Two cohorts are compared (Table 1). The pre–pandemic cohort includes students enrolled in Grade 8 in 2016/17 and Grade 10 in 2018/19, who thus made their track choice before the pandemic. The pandemic cohort includes students enrolled in Grade 8 in 2020/21 and Grade 10 in 2022/23, who chose their track during the pandemic period. INVALSI assessments were not administered in 2019/20 due to school closures.

Table 1. Description of cohorts.

Academic year	Pre–Covid cohort	Covid cohort
2022/23		Grade 10
2021/22		
2020/21		Grade 8
2019/20	No assessment	
2018/19	Grade 10	
2017/18		
2016/17	Grade 8	

Analytical strategy

The study uses a multilevel multinomial logistic model with school–level random intercepts, random slopes for the pandemic cohort, and provincial fixed effects. The outcome variable is upper–secondary track choice, and the primary predictor is the ESCS index for socioeconomic status (SES) of the student, divided into four quartiles: low, medium–low, medium–high, high. An interaction term between SES and cohort is included.

Four additive models are estimated. Model 0 is the null model, and contains only the predictor interacted with covid cohort, and province fixed effects. Model 1 includes controls for a baseline set of covariates including gender, migration background, nursery and kindergarten attendance, and on–time educational progression. Model 2 additionally controls for prior achievement using standardised INVALSI test scores in reading and mathematics, while Model 3 accounts for prior achievement using self–reported grades.

Preliminary results

Preliminary results of the four models are shown in Table 2.

Table 2. Multilevel multinomial logit models of upper–secondary track choice by socioeconomic status

	Technical *vs* academic track				Vocational *vs* academic track			
	M0	M1	M2	M3	M0	M1	M2	M3
Covid cohort (Ref: Pre–Covid)								
Covid cohort	–0.090*** (0.018)	–0.054** (0.019)	–0.130*** (0.020)	0.096*** (0.020)	–0.234*** (0.040)	–0.190*** (0.040)	–0.365*** (0.042)	0.056 (0.043)
Socioeconomic status (SES) (Ref: High SES)								
Low SES	1.482*** (0.016)	1.586*** (0.017)	1.355*** (0.018)	1.269*** (0.018)	2.660*** (0.030)	2.623*** (0.031)	2.209*** (0.032)	2.077*** (0.033)
Medium–low SES	1.077*** (0.016)	1.159*** (0.017)	1.013*** (0.017)	0.986*** (0.018)	1.786*** (0.031)	1.780*** (0.031)	1.519*** (0.033)	1.482*** (0.033)
Medium–high SES	0.747*** (0.016)	0.810*** (0.017)	0.704*** (0.017)	0.691*** (0.018)	1.187*** (0.032)	1.198*** (0.032)	0.998*** (0.034)	0.993*** (0.035)
Interactions (Covid cohort × SES)								
Covid × Low SES	0.169*** (0.022)	0.063** (0.023)	–0.143*** (0.024)	–0.088*** (0.025)	0.297*** (0.043)	0.232*** (0.043)	–0.171*** (0.046)	–0.001 (0.046)
Covid × Medium–low SES	0.137*** (0.022)	0.083*** (0.023)	–0.020 (0.024)	–0.015 (0.024)	0.165*** (0.044)	0.130** (0.044)	–0.055 (0.047)	–0.029 (0.047)
Covid × Medium–high SES	0.061** (0.022)	0.017 (0.023)	–0.037 (0.024)	–0.031 (0.024)	0.076 (0.046)	0.050 (0.046)	–0.038 (0.049)	–0.022 (0.049)
Individual controls	No	Yes	Yes	Yes	No	Yes	Yes	Yes
Achievement controls	No	No	WLE	Grades	No	No	WLE	Grades
Province fixed effects	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
School random effects	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
School clusters	5,548	5,548	5,548	5,548	5,548	5,548	5,548	5,548
N.	443,515	443,515	443,515	443,515	443,515	443,515	443,515	443,515

Individual controls include gender, migration background, nursery and kindergarten attendance, on–time school progression.

Standard errors in parentheses. * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$.

The four models reveal a consistent SES gradient in upper–secondary track choice, together with a reversal in the estimated COVID cohort effect depending on *which* achievement measure is used as a control. In the baseline model (M0) students in the lowest SES quartile are considerably more likely than their high–SES peers to enter technical or vocational tracks rather than the academic track, with the gap narrowing steadily across SES quartiles. The main COVID cohort effect is negative in M0, but the COVID × SES interactions are positive and significant for low and medium–low SES students, indicating that the pandemic cohort widened, rather than narrowed, the socioeconomic gap in non–academic track entry.

Adding individual–level controls (M1) leaves this pattern substantively unchanged; SES main effects for the technical track rise slightly, while those for the vocational track decrease marginally. The COVID × SES interactions decline uniformly in both tracks relative to M0.

The picture shifts markedly once achievement controls are introduced. When standardised test scores (WLE) are added (M2), the COVID main effect becomes more negative for both tracks, and, more importantly, the COVID × SES interaction for low–SES students reverses sign. Low–SES students in the pandemic cohort now appear less likely to enter technical (β = –0.14) and vocational (β = –0.17) tracks relative to their non–pandemic low–SES peers, once test–score performance is held constant. However,

when school grades are used as the achievement control (M3), the COVID main effect turns positive for the technical track (β = 0.10) and is no longer significant for the vocational track. The COVID × SES interaction for low–SES students in the technical track remains negative and significant but considerably decrease compared to M2, while for the vocational track it drops close to zero and loses significance altogether.

Discussion

The divergence between M2 and M3 is important, as it is not just a matter of degree but of direction for the COVID main effect: the same cohort coefficient is negative when achievement is accounted for using standardised test scores and turns positive (technical track) or null (vocational track) when using grades. This suggests that WLE scores and school grades are not interchangeable measures of "prior achievement" in this context, but capture different aspects of student performance. Standardised tests are external, blind to the pandemic disruption itself, and arguably provide a more stable benchmark of what students actually learned. Grades, by contrast, are assigned by teachers who were themselves navigating remote instruction, altered assessment practices, and documented tendencies toward leniency during school closures.

The COVID × SES interactions add a further nuance. In Model 2, low–SES students in the pandemic cohort appear markedly less likely to enter both technical and vocational tracks than their non–pandemic low–SES counterparts; in Model 3, this effect weakens rather than disappearing uniformly, as it shrinks by almost a third for the technical track, but decreases near zero and loses its significance for the vocational track. These findings may point to grade inflation, or compensatory grading, operating differently across the two tracks. If teachers inflated grades more for disadvantaged students, perhaps compensating for unequal home–learning conditions, controlling for grades would absorb part of the true achievement gap, attenuating or eliminating the residual interaction relative to the WLE specification.

This interpretation is consistent with a broader distinction between primary and secondary effects (Boudon, 1974): the WLE–based specification (M2) more plausibly isolates a residual secondary effect (a shift in track choice net of actual learning) while the grade–based specification (M3) partly re–absorbs part of that same secondary effect into the achievement control itself, since grades are not immune to the same institutional and relational dynamics that shape tracking recommendations. In other words, using grades as a proxy for achievement may understate, rather than eliminate, the true extent to which the pandemic reshaped track choice independently of learning, precisely because grades may themselves have been affected by the pandemic in ways correlated with SES.

All in all, these findings call for caution in using "achievement controls", particularly for pandemic–era cohorts, and suggest that the sensitivity of results to the choice of achievement measure may itself vary across tracks.

Keywords: Track choice; Socioeconomic status; COVID–19; INVALSI–SNV

Evolving Equity in Finnish PISA Results – Insights from Low–Performing Students

Jonna Pulkkinen – Heli Kauppinen – Arto K. Ahonen

Introduction

In international comparison, Finland has long been a model country for good learning outcomes (Simola et al., 2017) as well as for success in reducing educational inequalities (Hadjar & Uusitalo, 2016). Low inequalities have been explained, for example, by a social–democratic welfare regime with a low stratified education system (Hadjar & Uusitalo, 2016; Lasso De La Vega et al., 2020). The Finnish educational policies have placed strong emphasis on achieving educational equity (Chong, 2018). Instead of central steering, performance accountability, and high–stakes assessment, the Finnish education model has been characterised by decentralisation, professional trust, and school–based assessment. Thus, municipalities and schools, as well as teachers, have a high degree of autonomy. Despite this, differences between schools have mainly been small in Finland (Chong, 2018).

The strength of the Finnish education system has been considered to be the fact that differences due to students' backgrounds have been able to be evened out and that the share of low–performing students has been relatively low. Recent results of large–scale assessments (e.g., PISA study) show that the performance of Finnish students has declined and the share of low–performing students has increased (OECD, 2023). Between school variation has remained among the least among all the participating countries and economies and was 9 percent in PISA 2022, where the OECD average was 32 per cent (OECD, 2023). Still,

there has been a slight increase also in the between school variation and differences in performance between schools (Ahonen, 2021).

Although Finland is known for educational equality, there have been clear differences in performance between groups that are based on students' background factors. According to international large–scale assessments (ILSAs), the differences in reading between girls and boys have been quite large (e.g., Mullis et al., 2023; OECD, 2023). Instead, in mathematics, there have been no similar differences and boys' and girls' performance has been mainly on the same level (von Davier, 2024; OECD, 2023). The performance gap between native and immigrant students is very large in Finland, even if this gap has narrowed recently according to PISA 2022 results (OECD, 2023). Partly the large performance gap between native and immigrant students in Finland is explained by differences in socio–economic status (SES) that is lower for immigrant students than for native students. In terms of SES, Finland has traditionally succeeded to equalize the differences in performance between low– and high–SES students. Although the SES gap has increased somewhat in recent years, Finland still succeeds in equalizing this difference better than OECD countries on average (OECD, 2023).

In this study, we examine the trends in educational equality in Finland using PISA data. We focus on the low–performing students and study changes in educational equality among these students. The research questions are: (1) How do changes in educational equality manifest themselves in the group of low–performing students from 2003 to 2022?, (2) Which factors do explain low performance in mathematics in 2012 and 2022?

Data and Methodology

In this study, we used the Finnish PISA data from years 2003, 2006, 2009, 2012, 2015, 2018 and 2022. The number of assessed students ranged from 4,714 (in PISA 2006) to 10,239 (in PISA 2022). In this study, we focus on students' performance in mathematics that was the main subject of PISA in 2003, 2012 and 2022. In PISA 2012 and PISA 2022, students with immigrant background were oversampled in Finland, which provides opportunities to draw more reliable conclusions regarding students with an immigrant background. Since the research questions also examine proficiency differences by immigrant background, the latter research question is limited to the PISA cycles with oversampling. In addition, only PISA 2012 and 2022 data are used in the analysis of the first research question with regard to immigrant students.

In PISA, students' performance in mathematics can be divided into performance levels based on the mathematics score points. In this study, we focus on students that perform below Level 2 (i.e., low–performing students). According to OECD (2023, p. 89), "Level 2 is considered the baseline level of proficiency students need to participate fully in society". We also used four different background variables by which we studied educational equality: gender, socio–economic status (SES) that was based on the ESCS index of PISA, immigrant background, and the community location of school.

All statistical analyses within this study were conducted with R software and its packages intsvy and Rrepest. Both packages are designed for working with large–scale international educational assessment datasets. In all analyses, the complex survey design was accounted for by using proper sampling and replicate weights. In addition, in analysis considering mathematics achievement, all plausible values were used. Until 2015, PISA studies used five plausible values, and since 2015, ten plausible values have been used.

To address the first research question, the changes in proportions of low–performing students in mathematics were examined using descriptive statistics. The statistical significance of changes between PISA cycles and differences between groups were statistically tested using t–test. To answer to the second research question, the logistic regression model was fitted to Finnish PISA datasets of 2012 and 2022. Prior to logistic regression, plausible values were recoded as dichotomic variables indicating whether student is low–performing or not (1 = achievement lower than proficiency level 2; 0 = achievement higher than proficiency level 2). This recoded variable was used as the outcome variable in the logistic regression model. Students' gender, SES (ESCS index), immigration background, and school location were included as explanatory variables. The relationships between the explanatory variables and the outcome variable are interpreted using odds ratios (OR), which describe the odds of being a low–performing student in mathematics relative to the reference group in the model.

Results and discussion

In the 2000s when Finland was at the top of the PISA study, the share of low performers was relatively low (6–8%). However, the share of low–performing students has increased since 2009. In 2022, the share was as high as 25%, which was statistically significantly higher than in all previous cycles ($p < .001$). This shows

that, in addition to the decrease in the average level of performance, the decrease in the performance of the Finnish students is also reflected in a significant increase in the share of low–performing students.
When studying changes in the share of low–performing students in different groups, it can be noticed that educational inequalities have also increased as performance level has declined in Finland. Since 2012, when students' performance in mathematics turned to a clear decline in Finland, the shares of low–performing students between girls and boys also began to differentiate. In 2012, the share was 10% for girls and 14% for boys, and for the first time, the difference in the share of low–performing students was statistically significant ($p < .001$). The differentiation of performance can also be seen when looking at the share of low–performing students in the national quartiles of SES. In each PISA cycle, there has been a statistically significant difference ($p < .001$) in the shares of low–performing students between lowest and highest quartile of SES. However, the difference has increased since 2009, as the share has grown more among students in the lowest quartile. In 2012, the gap had grown to 15 percentage points, and in 2022, the difference between lowest and highest quartile was as much as 28 percentage points, which was statistically significantly higher compared to the previous cycles ($p < .001$). More than half of the first–generation immigrant students were low performers both in 2012 and 2022, whereas the share of low–performing native students was 10% in 2012 and 22% in 2022. However, the difference between native students and students with an immigrant background has levelled off slightly in ten years, which is explained by the fact that the share of the low–performing native students has increased more. Despite this, the results do not show positive development in education equality, as the share of low–performing immigrant students remained high in Finland and even increased between 2012 and 2022. The fourth variable by which we examined the realisation of educational equality was the community the school is located. Differences in the share of low–performing students have generally been small in Finland up to 2022, when the difference (8 percentage points) between schools located in cities and schools located in villages was for the first time statistically significant ($p < .01$). The results of logistic regression supported the descriptive results. When all four explanatory variables were in the same logistic model, boys ($OR = 1.32$, $p < .01$) and students with an immigrant background (first generation: $OR = 6.21$, $p < .001$; second generation: $OR = 3.81$, $p < 001$) were more likely low–performing students, whereas students with higher SES ($OR = 0.51$, $p < .001$) performed less likely low in 2012. In 2022, the results of logistic regression were mainly similar. However, at score level, the odds ratios for immigrant background students were slightly smaller in 2022 than in 2012. This is primary because native students' overall performance had a stronger decline compared to the peers with immigrant background. In addition, unlike the results of 2012, the results for 2022 indicated that, compared to students in rural schools, students in city schools were less likely ($OR = 0.73$, $p < .05$) low–performing students in mathematics. All in all, the examination of trends in the share of low–performing students in different groups indicates growing educational inequalities in Finland. In the 21st century, urbanisation has increased in Finland. Especially in large cities, some factors have been identified that increase educational inequalities. These are, for example, housing policies, urban spatial segregation, and school choice that may have unintended consequences of reinforcing segregation and excluding students with lower SES and immigrant backgrounds from certain schools (Smeplass et al., 2024). Moreover, some organisational models, such as differentiated support models for students with special educational needs or immigrant background, that are intended to promote equity may also exclude these students from the general education system and thus cause educational inequalities (Smeplass et al., 2024).

Conclusions

Finland has long been successful in international comparisons of learning outcomes and has been known for its equity. On the first decade of 21st century Finnish basic education system was able to even out the effect of the low socio–economic background more than the OECD countries on average (Ahonen, 2021). This meant that the Finnish school system was able to produce good learning outcomes for vulnerable students as well. This situation seems to have weakened somewhat by the 2010s, and reached OECD average in 2015, raising concerns, for example, among policymakers. Many recent educational policy reforms have been justified by the decline in learning outcomes. In addition, recent governments have sought to improve the learning outcomes of Finnish students, for example, by strengthening educational equality and equity, and by reforming the support system for learning and schooling. Thus, many policy actions have been or will be taken to improve learning outcomes and increase educational equality, as well as support students who are vulnerable. These actions are needed to reverse the development of learning outcomes and educational equality in Finland.

Keywords: Equality; Learning outcomes; Socioeconomic status; Immigrant background

Occupational Expectations in European Educational Regimes: An Application of the ESeC Indicator to OECD–PISA 2022 Data

Eleonora Mattarelli – Pasquale Di Padova – Orazio Giancola – Enrico Nerli Ballati

Introduction

The formation of occupational expectations in adolescence plays a role in the reproduction of social stratification. Career plans are not mere individual preferences; rather, they represent structured ambitions shaped by family background, institutional context, and academic performance. The sociological literature on status attainment distinguishes between primary effects, (operating through actual academic achievement) and secondary effects (shaping educational and occupational choices independently of performance) (Boudon, 1974; Breen & Goldthorpe, 1997). A third category, tertiary effects of social origin, further operates within educational tracks and levels, reproducing social inequalities even among students in comparable institutional positions. All three mechanisms are active well before students enter the labour market.

Cross–national comparative evidence, including analyses based on PISA cycles, has established that these mechanisms vary considerably across institutional settings. Sikora and Saha (2007) showed that country–level structural inequality is paradoxically associated with higher occupational expectations among 15–year–olds, even after controlling for family background and school performance. Furthermore, Sikora and Pokropek (2011) documented systematic gender differences, showing that while girls generally project themselves into higher–status occupations than boys, horizontal occupational segregation strongly shapes choices within status categories.

West and Nikolai (2013) conceptualize these institutional boundaries through the lens of "education regimes", identifying four distinct clusters in Europe. The Nordic regime is characterized by comprehensive school structures, low early tracking, high public expenditure, and strong emphasis on equality of opportunity. The Continental regime combines early and rigid tracking with a well–institutionalized vocational pathway, tightening the link between education and labour market destinations. The Mediterranean regime displays lower public expenditure, late formal tracking with high internal segmentation, and a weaker institutional coupling between schooling and employment. The English–speaking regime is marked by private provision, selective elements within formally comprehensive structures, and market–oriented features. Each regime configuration generates specific patterns of educational inequality and is associated with distinct expenditure profiles.

Recent national evidence has applied the European Socio–economic Classification (ESeC) (Rose & Harrison, 2010) to PISA 2018 data for Italy, establishing a rigorous baseline (Nerli Ballati & Palmerio, 2024). Results revealed a pronounced upward inflation of occupational expectations: nearly half of Italian 15–year–olds project themselves into the higher salariat (ESeC Class 1 and 2) and confirmed that secondary effects persist even when controlling for PISA competence scores, creating a structural mismatch with the actual domestic labour market.

This study extends that methodological framework to a fully comparative cross–national design. It applies the simplified ESeC 2008 classification to PISA 2022 data across all participating European countries, placing a specific analytical focus on four countries selected as archetypes of the educational regimes identified: Germany (Continental), Finland (Nordic), Italy (Mediterranean), and the United Kingdom (English–speaking).

Data

The study uses public microdata from the OECD Programme for International Student Assessment (PISA) 2022 cycle.

Three variables from the student questionnaire serve as primary inputs. OCOD1 contains the 4–digit ISCO–08 code of the mother's occupation, derived from an open–ended text response officially coded by the OECD. OCOD2 contains the equivalent code for the father's occupation. OCOD3 contains the 4–digit ISCO–08 code of the occupation the student expects to hold at age 30, coded from the same open–ended format. The parental class is determined via a dominance approach, assigning the student to the higher ESeC class of the two parents.

Supplementary variables include: proficiency levels in reading, mathematics, and science, gender, and migration background.

Methodology
The core methodological operation involves mapping the 4–digit ISCO–08 codes for OCOD1, OCOD2, and OCOD3 to the simplified version of the European Socio–economic Classification (ESeC 2008) to derive classes for maternal occupation, paternal occupation, and student expected occupation. The same approach was adopted by Nerli Ballati and Palmerio (2024) for the Italian case.
The analytical core consists of intergenerational mobility tables, constructed separately for each country, crossing parental ESeC class of origin with student expected ESeC class of destination. While the broader cross–national comparison identifies macro–level patterns, detailed interpretation focuses on the four archetype countries. To isolate the net contribution of academic performance and social origin, the mobility tables are stratified by PISA proficiency levels and socio–demographic variables.
Expected Results and Discussion
As the study is currently in its analytical phase, the following main results are hypothesized based on the underlying theoretical framework.
First, across all European countries, the outflow distributions of the mobility tables are expected to display a pronounced upward shift, with most students projecting themselves into ESeC Classes 1 e 2. This pattern reflects a widespread inflation of adolescent occupational aspirations, already documented in the Italian case. However, the magnitude of the mismatch between these expectations and the actual labour market structure will vary significantly by education regime.
Second, the diagonal cells of the mobility matrices, indicating direct intergenerational class reproduction in expectations, are expected to be most pronounced in Germany (due to early and rigid tracking) and Italy (due to high internal segmentation).
Third, reflecting the egalitarian and comprehensive structure of the Nordic regime, Finnish mobility matrices are expected to show the weakest link between parental class of origin and student expected destination class. Conversely, the UK matrices are expected to show high absolute aspiration levels across all class backgrounds, paired with strong horizontal differentiation within destination categories, particularly by gender.
Fourth, stratification of the mobility tables by PISA proficiency levels is expected to confirm that within comparable performance brackets the association between social origin and expected destination remains statistically meaningful in Germany, Italy, and the UK, while Finland is expected to display the smallest net residual effect.
Conclusions
This study advances the methodological integration of official statistics by systematically applying standardized socio–economic classifications directly to large–scale international educational assessments. By shifting the focus to a fully comparative design, the mobility table approach enables a direct cross–national comparison of how different educational regimes shape the origin–destination structures embedded in adolescent expectations. Ultimately, these expected results will provide robust empirical grounding for timely policy discussions regarding the alignment of training supply, school–to–work transition pathways, and the reduction of social inequalities across Europe.

Keywords: Occupational Expectations; ESeC; PISA 2022; Educational outcomes

Migration Background and Digital Inequalities Among Adolescents in Spain

Diana Fernández Méndez– José Manuel Amoedo Meijide – Jackiline Salazar Torres

Introduction
Students with a migration background consistently face educational disadvantages compared to their native peers across Europe (Rodrigues, 2018; van Dijk & Hacker, 2003). These disparities are reflected not only in academic achievement but also in the opportunities to benefit from digital technologies for learning. While digitalisation has the potential to promote educational inclusion, it may also reinforce existing inequalities when access to digital resources, digital skills, or educational uses of technology differ across social groups (Akinlar et al., 2023).
Previous research has shown that immigrant students generally begin using digital technologies later than native students but tend to report more intensive ICT use during adolescence. Moreover, ICT–related factors explain part of the educational achievement gap between immigrant and native students, suggesting

that digital policies may contribute to reducing educational inequalities. Other students have highlighted that educational ICT use outside school varies according to migration background, socioeconomic status, and gender, with the dimensions interacting rather than operating independently (Becker, 2021). At the same time, recent literature reviews have emphasized that the rapid digitalisation of education, particularly following the COVID–19 pandemic, has disproportionately affected migrant and refugee students, exposing persistent inequalities in digital access, digital skills, and educational opportunities (Akinlar et al., 2017).
However, existing evidence has largely examined access to ICT, digital competences, or ICT use separately (Martínez–Gautier et al, 2021). This fragmented approach overlooks the multidimensional nature of digital inequalities and may underestimate the cumulative disadvantages experienced by students with a migration background. Moreover, little attention has been paid to distinguishing between first– and second–generation immigrant students or to estimating whether observed differences persist after accounting for socioeconomic and educational characteristics.
To address these gaps, this study adopts a multidimensional perspective of digital vulnerability by jointly analysing four complementary dimensions: access to digital resources at home, access at school, digital competences, and ICT use. Using data from PISA 2022 for Spain and Galicia, we compare native students with first– and second–generation immigrant students and estimate both gross and net digital gaps through Propensity Score Matching, allowing us to isolate the effect of migration background on digital vulnerability.

Data and Methods

To measure vulnerability in access to ICT at home (DVHOME) and at school (DVSCHOOL), the synthetic indicators ICTHOME and ICTSCH developed by PISA are used. ICTHOME measures the availability and quality of digital resources at home, including digital devices and resources such as desktop and laptop computers, mobile phones, e–book readers, tablets, educational software, Internet access, and televisions. ICTSCH measures the availability and quality of digital resources at school, including similar aspects to those considered for households.
Digital skills vulnerability (DVSKILLS) is measured using the synthetic indicator ICTEFFIC, which captures students' self–reported digital skills across a range of competencies. The 16 items included assess abilities such as using word processors, searching for information online, evaluating the reliability of online information, protecting personal privacy, and programming algorithms to solve specific problems. Since digital skills are based on students' self–assessments, which may be subject to important biases, a correction factor is proposed to reduce these biases. This correction factor is constructed using the normalized average of the mathematics plausible values (MATHEMATICS), ranging from zero to one, and students' normalized self–perceived mathematics ability (MATHEFF), also ranging from zero to one, following Equation (1).

$$CORRECTOR = 1 + (MATHEFF - MATHEMATICS) \quad (1)$$

Consequently, DVSKILLS is calculated according to Equation (2).

$$DVSKILLS = ICTEFFIC * CORRECTOR \quad (2)$$

To measure vulnerability related to ICT use (DVUSE), PISA does not provide a synthetic indicator. Therefore, this study proposes three indicators based on the main available information regarding ICT use both at school and at home, presented in Table 1.

Table 1. ICT use variables

Variable	Description
LEARNSCHOOL	Hours per day that students spend using ICT for learning activities at school
LEARNBA	Hours per day that students spend using ICT for learning activities before or after school
LEARNWK	Hours per day that students spend using ICT for learning activities during the weekend
LEISURESCHOOL	Hours per day that students spend using ICT for leisure activities at school
LEISUREBA	Hours per day that students spend using ICT for leisure activities before or after school
LEISUREWK	Hours per day that students spend using ICT for leisure activities during the weekend

Source: Authors' elaboration based on OECD (2023).

To account for the diminishing positive effects of ICT use for learning and the increasing negative effects of ICT use for leisure, the indicator DVUSE is defined according to Equation (3).
The parameter α takes strictly positive values, so that higher values imply greater sensitivity to leisure–related ICT use. To capture different degrees of sensitivity, α is set equal to 1, 1.5, and 2, yielding the indicators DVUSE1, DVUSE1.5, and DVUSE2, respectively.
Based on these indicators, the vulnerability variables IDVHOME, IDVSCHOOL, IDVSKILLS, DVUSE1, DVUSE1.5, and DVUSE2 are constructed, classifying students into three levels of digital vulnerability:

- No vulnerability: the individual's value is above 80% of the sample median.
- At risk of vulnerability: the individual's value is between 60% and 80% of the sample median.
- Vulnerable: the individual's value is below 60% of the sample median.

Using these four dimensions, three multidimensional digital vulnerability indicators (DV1, DV1.5, and DV2) are constructed, one for each ICT use specification. These indicators are based on a multidimensional conception of vulnerability in which dimensions cannot be fully compensated by one another. The material vulnerability dimensions (DVHOME, DVSCHOOL, DVSKILLS, and DVUSE) are summed and students are classified into three categories:

- No vulnerability: total score less than or equal to two.
- At risk of vulnerability: total score between zero and two.
- Vulnerable: total score above six.

Gross digital divides are estimated by calculating the incidence of vulnerability among native students (NATIVE) and students with a migration background (IMMIGRANT), both for the overall vulnerability indicators and for each individual dimension. The analysis further distinguishes between first–generation (FIRST GENERATION) and second–generation (SECOND GENERATION) immigrant students to assess whether digital vulnerability differs by immigrant generation.
Net digital divides are estimated using several Propensity Score Matching (PSM) techniques to isolate the effect of migration background on digital vulnerability. The treatment and control groups used in the matching procedures are shown in Table 2.

Table 2. Treatment and control groups in the matching procedures

Treatment	Control
IMMIGRANT	NATIVE
FIRST GENERATION	NATIVE
SECOND GENERATION	NATIVE

Source: Authors' elaboration based on OECD (2023).

To achieve well–balanced matches, several covariables are defined. Exact matching is combined with Nearest Neighbour Matching (NNM) using K = 1, 3, 5, and 10 neighbours, and a caliper of 0.5 for K = 3, 5, and 10. Balance before and after matching is assessed through the standardized mean bias of the propensity score and individual covariates, the pseudo–R^2, and graphical inspection of the propensity score distribution. To ensure the most robust comparison possible, preference is given to matching specifications that simultaneously achieve good balance and retain the largest possible control sample.

Results and Discussion

able 4 reports the incidence of digital vulnerability before matching. The results reveal significant gross digital divides between native and immigrant students across all three multidimensional indicators (DV1, DV1.5 and DV2), with immigrant students consistently showing higher levels of digital vulnerability. These differences are mainly explained by inequalities in access to digital resources at home (DVHOME) and ICT use (DVUSE), whereas differences in access at school (DVSCH) and digital skills (DVSKILLS) are comparatively smaller.

Table 4. Digital vulnerability in native and immigrants' students prematching

		Vulnerable (%)	At risk (%)	Not vulnerable (%)	Rao–Scott adjusted chi–square test
DV1	Native	22.18	21.69	56.14	8.384****
	Immigrant	27.09	20.92	51.98	
DV1.5	Native	19.53	20.08	60.40	7.705****
	Immigrant	23.91	20.16	55.93	
DV2	Native	17.29	19.19	63.51	9.343****
	Immigrant	21.82	19.74	58.44	
DVHOME	Native	3.14	8.93	87.93	14.816****
	Immigrant	4.31	13.09	82.61	
DVSCH	Native	7.88	11.76	80.37	2.307*
	Immigrant	9.54	10.99	79.47	
DVSKILLS	Native	0.87	3.20	95.93	3.411**
	Immigrant	0.50	4.20	95.31	
DVUSE1	Native	8.01	13.83	78.16	3.843**
	Immigrant	10.16	12.91	76.93	
DVUSE1.5	Native	5.22	10.71	84.07	3.031**
	Immigrant	6.85	10.58	82.58	
DVUSE2	Native	3.03	8.49	88.49	7.314****
	Immigrant	5.01	8.12	86.87	

Source: Authors' elaboration based on OECD (2023).Table 5 presents the results after applying Propensity Score Matching. Once differences in socioeconomic, demographic, and educational characteristics are taken into account, the multidimensional digital vulnerability gaps between native and immigrant students are no longer statistically significant. Similarly, no significant differences remain for vulnerability related to ICT use, digital skills, or access to digital resources at school. The only remaining statistically significant difference is observed in access to digital resources at home (DVHOME), indicating that inequalities in household digital resources persist even after controlling for observable characteristics.

Table 5. Digital vulnerability in native and immigrants' students postmatching

		Vulnerable (%)	At risk (%)	Not vulnerable (%)	Rao–Scott adjusted chi–square test
DV4.3.1	Native	23.89	20.71	55.40	1.081
	Immigrant	25.26	21.29	53.46	
DV4.3.15	Native	21.28	19.41	59.31	1.037
	Immigrant	21.92	20.65	57.43	
DV4.3.2	Native	18.88	18.57	62.54	1.954
	Immigrant	19.90	20.17	59.93	
DVHOME	Native	3.71	9.78	86.51	4.816**
	Immigrant	3.89	12.39	83.72	
DVSCH	Native	8.25	10.39	81.36	0.329
	Immigrant	8.54	10.96	80.50	
DVSKILLS	Native	0.96	4.02	95.02	0.582
	Immigrant	0.71	3.81	95.47	
DVUSE1	Native	9.25	13.32	77.42	0.246
	Immigrant	9.25	12.67	78.08	
DVUSE1.5	Native	6.22	10.96	82.82	0.557
	Immigrant	5.96	10.13	83.92	
DVUSE2	Native	3.85	8.50	87.66	0.005
	Immigrant	3.85	8.58	87.57	

Source: Authors' elaboration based on OECD (2023).

These aggregate results indicate that a considerable share of the digital divide associated with migration background is explained by compositional differences between native and immigrant students. The following analysis examines whether this pattern differs between first– and second–generation immigrant students.

Conclusions

This study provides a multidimensional assessment of digital vulnerability among adolescents with and without a migration background in Spain and Galicia. The results show that immigrant students exhibit higher levels of digital vulnerability before controlling for socioeconomic and educational characteristics. However, most of these differences disappear after Propensity Score Matching, suggesting that the observed digital divide is largely explained by differences in students' socioeconomic profiles rather than migration background itself. Nevertheless, inequalities in access to digital resources at home persist, highlighting the importance of policies aimed at reducing material digital deprivation among immigrant households. The subsequent analysis by immigrant generation further reveals that these patterns are not homogeneous, underlining the need for more targeted policy interventions.

Acknowledgements

The authors would like to thank the funding provided through the "Proxecto colaborativo 2026: Fendas dixitais entre adolescentes con e sen orixe migratoria en España e Galicia: implicacións para o fracaso escolar" by the IDEGA.

Keywords: Educational inequality; Territorial disparities; Learning outcomes

Who Matters for Whom? Non–linear Peer Effects on Achievement and Absenteeism

Iván Vicente – Tommaso Agasisti – Melisa L. Diaz Lema

This paper examines the causal impact of peer quality on academic achievement and school absenteeism among Italian middle school students, exploiting the quasi–random reassignment of peers at the primary–to–middle school transition. Using administrative data on approximately 257,000 students linked across grades 5 and 8, we identify peer effects from variation in the prior achievement of newly assigned classmates, measured before students meet, relative to the set of peers they shared in primary school. For cognitive outcomes, we employ both a between–student model with school–transition fixed effects and a within–student stacked design that exploits variation across subjects within the same student, absorbing all time–invariant individual characteristics. For absenteeism, we use a cross–sectional model with school–transition fixed effects, complemented by entropy balancing to address selection on observables. Average peer effects are positive and significant for achievement — a one standard deviation increase in new peers' prior scores raises grade 8 scores by 0.013–0.047 standard deviations — and negative for absenteeism, with higher–quality peers reducing absences by 0.017–0.023 standard deviations. Our central contribution is documenting the non–linear, heterogeneous nature of these effects. For academic achievement, the benefits of higher–quality peers follow an inverted U–shape across the own prior score distribution: effects are negligible for very low performers, peak for students around the middle of the distribution, and fade for high achievers. For absenteeism, the pattern differs notably, the attendance benefits of better peers are strongest among weaker students and monotonically decline with own prior performance. This nuanced asymmetry across outcomes suggests that peer influence operates through related but distinct channels for cognitive and behavioral dimensions of human capital, with implications for how classroom composition policies may differentially affect students depending on the outcome of interest.

Keywords: Peer effects; Academic achievement, Absenteeism, School transitions

AUTHORS

Giovanni Abbiati is an associate professor of Sociology at the University of Brescia's Department of Economics and Management. His research focuses on social inequalities in education and public policy evaluation, with particular emphasis on field experiments in educational settings. His work has appeared in international peer-reviewed journals, and he contributes regularly to the Italian policy debate through Lavoce.info.

Tommaso Agasisti He teaches Public Management in the Master of Science in Management Engineering and Economics and Business Organization at the Bachelor in Production Engineering, Lecco Campus of PoliMi. Since 2024, he has been Associate Dean for Institutions and Public Administrations (IPA) at PoliMi GSOM - Graduate School of Management (where he served as Associate Dean for International Relations and Quality from 2020 to 2023). Since 2023, he is the Rector's Delegate for Learning Analytics. From 2019 to 2024, he was a member of the Board of Governors of Politecnico di Milano.

Arto K. Ahonen, Ph.D. in Educational Sciences, works as a university researcher at the Finnish Institute for Educational Research at the University of Jyväskylä. He has served as the national project manager for the PISA study since 2016, overseeing its implementation and reporting for the 2018, 2022, and 2025 rounds. Ahonen has the title of docent (educational evaluation) at the University of Tampere.

Fabio Aiello is an Associate Professor of Social Statistics at UKE. His main research interests include biostatistics and meta-analysis, statistical methods for constructing composite indicators, Rasch models for questionnaire validation, and the development of composite indicators.

Carmen Aina is Associate Professor of Economic Policy at the Department for Sustainable Development and Ecological Transition, University of Piemonte Orientale. Her research focuses on the economics of education, labour economics, inequality, and policy evaluation. She also serves as an expert for the Italian Ministry of University and Research on analysis and evaluation of the national university system.

Fleana Albertoni is a Technology teacher at I.C. "Dante Alighieri" in Aulla. She also serves as the Vice-Principal and First Collaborator to the School Principal within the institutional governance team.

Alexander Álvarez-González holds a PhD in Psychology and is a researcher specializing in behavioral and health sciences methodology. Experienced in evaluating clinical significance and treatment outcomes, his work focuses on psychometrics, questionnaire validation, and statistical analysis and gifted education assessment.

Daniel April is an Associate Research and Policy Analyst with UNESCO's Global Education Monitoring Report, where he has worked since 2018. He coordinates the Profiles Enhancing Education Reviews (PEER). His research focuses on teachers and school principals. He holds a PhD in Educational Administration and Policy Studies from Université Laval and previously served as an Academic Director and French as a Second Language teacher in Canada.

Arantza Arruti holds a PhD in Pedagogy and an Expert in Leisure Education from the University of Deusto (UD). She is: tenured Professor in the Department of Education (Faculty of Education and Sport-UD), where she is since 1996; EntreComp Champion (European Commission); head of Projects and Conferences on Teaching Innovation (Teaching Innovation Unit); member of the committee for the accreditation and evaluation of teaching quality; member of eDucaR (Deusto Education Research) research team. Her research focuses on curriculum design, development, and innovation; active learning methodologies; and entrepreneurial competence.

Massimo Attanasio is Full Professor of Social Statistics at the University of Palermo since 2002, where he teaches Social Statistics and Biostatistics. His research interests focus on university student mobility, gender studies, and the applications of statistics in medicine and public health.

Dragana Avramov is the Director and senior scientific fellow of Population and Social Policy Consultants. Since 1999 she has been working as independent expert for the European Commission on evaluation of project proposals, policy review, and as Independent Observer for EC. Her research focuses on the social dimensions of education, managing migration and integration of immigrants, social impact, and ethics.

Patricia Ayllón Salas is a predoctoral researcher (FPU) in Educational Sciences at the University of Granada, Department of Developmental and Educational Psychology. Her research focuses on non-cognitive skills, academic performance, self-regulated learning, and service-learning, with particular attention to educational interventions aimed at improving outcomes for students in diverse and socially disadvantaged contexts.

Davide Azzolini is researcher at FBK-IRVAPP and Affiliated Scholar at the Urban Institute. His research focuses on educational inequality, student assessment, and policy-relevant evidence in education. He holds a PhD in Sociology and Social Research.

Barbara Baldazzi graduated in Demographic and Social Statistical Sciences and has a Doctorate degree in Research Methods for Analyzing Socioeconomic Change. She has been in Istat since 1997 on on social statistics, well-being, poverty and education.

Paolo Barabanti is a researcher at INVALSI (Italian National Institute for the Evaluation of the Education System), where he works on large-scale educational assessment, learning outcomes, and educational data analysis. His research interests include academic resilience, gifted and high-achieving students, digital competences, artificial intelligence in education, and educational evaluation. He also teaches research methodology and educational assessment at university level.

Sabrina Lucilla Barone is a PhD student in Equity, Diversity and Inclusion at the Pegaso Telematic University and a teacher of literary subjects in secondary school. She is the author of scientific contributions dedicated to the pedagogy of artificial intelligence, Universal Design for Learning, inclusion, teaching professionalism and educational processes in school contexts.

Piera Bello is Associate Professor of Economics at the the Department of Economics of the University of Bergamo and Research Associate at ZEW-Leibniz Centre for European Economic Research. She is an applied microeconomist mainly interested in Public, Gender and Pension economics.

Silvia Bennati is a veteran school administrator currently serving as the School Principal for both the IPSSEOA "G. Minuto" in Massa and the I.C. "Dante Alighieri" in Aulla.

Maria Berrittella is Associate Professor in Public Finance and Economics at the University of Palermo, Department of Economics, Business and Statistics. Her current research interests include education mobility, inequality and policy evaluation.

Ivan Blancato PhD candidate at the Department of Sociology and Social Research, University of Milan–Bicocca. Research interests include education systems and the teaching profession, with a focus on precarious employment. Participation in several research projects on schools and education policies. Currently involved in the Quarta Indagine sulle Condizioni di Vita e di Lavoro degli Insegnanti Italiani.

Enrico Bocciolesi is Associate Professor of General and Social Pedagogy at the University of Urbino Carlo Bo, Department of Humanities. His research focuses on pedagogy, literacy, critical and complex thinking, learning styles, and lifelong education. He has authored numerous scientific publications and contributes to international research on education, inclusion, and teacher training.

Luca Bonacini is an Assistant Professor of Economics at the University of Eastern Piedmont and a Research Fellow at the Global Labour Organization (GLO). His research lies in applied microeconomics, with a focus on income distribution, inequality of opportunity, and labor market dynamics, particularly in relation to education and gender disparities.

Daniele Bondonio (Ph.D., Carnegie Mellon University) is a Full Professor of Economic Statistics at the University of Piemonte Orientale. He has served as a visiting scholar at UC Irvine, The Ohio State University, and the University of Toronto. Specializing in counterfactual impact evaluation, his research uses quasi-experimental methods to evaluate public policies in SME support, ICT, innovation, education, EU structural funds.

Stefano Bonometti is Associate Professor of Education in the Department of Human Sciences for Territorial Innovation (DiSUIT) at the University of Insubria, Italy. He chairs the degree programme in Languages and Competences for Education and Training and is a member of the Scientific Board of the University's Teaching and Learning Centre. His research is grounded in Cultural-Historical Activity Theory, with a particular focus on the Change Laboratory as an interventionist research method in organisational, healthcare and community settings. His work also addresses e-learning and faculty development.

Lidia Bordanoba is a Joan Oró FI Pre-doctoral Researcher in the Department of Sociology at the University of Barcelona. Her research focuses on the analysis and application of scientific evidence in the fields of education and gender, with a particul ar emphasis on its contribution to social improvement.

Raffaello Bronzini Director at the Bank of Italy, currently serving in the Structural Economic Analysis Directorate of the DG Economics, Statistics and Research. He previously headed the Research Unit at the Ministry of University and Research and the Research Division at the Bank of Italy's Rome Branch. He serves on several scientific and academic committees. His publications focus on innovation policy, regional development, and public research.

Giorgio Brunello is Professor Emeritus of Economics in the Department of Economics and Management of the University of Padua. He is an applied economist, with research interests in education, health, and labour economics.

Elsa Maria Bruni is Rector of IUL University and member of the INDIRE Board of Directors. Her research focuses on pedagogical theory, teacher education, inclusive processes, educational care and the epistemological foundations of education, with particular attention to the relationship between education, dignity and human development.

Luca Bungaro holds a PhD in Statistics (SECS-S/01) and has been working at INVALSI since 2026 within the Methodology and Psychometrics Unit. His research focuses on Item Response Theory (IRT), computer-based testing (CBT), computerized adaptive testing (CAT), response-time (RT) analysis in large-scale assessments, and psychometric models applied to educational assessment, particularly in the Italian context.

Claudia Califano is a teacher of Literary Subjects and Latin in High Schools. She has collaborated with Indire and Invalsi on the development of national projects, and as an expert e-tutor she participated in the ValeS Project and the Evaluation and Improvement Project. She is listed among Indire's experts for improvement and has carried out teacher training activities focused on didactic innovation and new technologies.

Marta Campagnoli is a PhD student at the University of Milan and is affiliated with the Milan Higher Education Observatory (MHEO). Her work focuses on data science applications in higher education, with particular attention to the analysis of administrative education data. She has also gained research experience at the Italian Ministry of University and Research.

Rina Camporese graduated in Demography and has a Doctorate degree in New Technologies for Information on Territory and Environment. She has been in Istat since 1994 on social statistics, data dissemination, statistical literacy and well-being.

Emiliano Campodifiori, is a statistician with a degree in Statistical and Economic Sciences from Sapienza University of Rome. He is a Technologist at the INVALSI Statistical Department, where he specializes in the analysis and transformation of large-scale datasets from the national assessments of student learning. His expertise focuses on quantitative methodologies applied to complex databases.

Francesca Carimando is a PhD candidate in sociology, at the University of Bologna. Her dissertation focuses on the impact of the COVID-19 pandemic on ethnic inequalities in education in Italy, including both disruption and recovery phases. Beyond her core research, she is interested in social stratification, including ethnic, gender, and socioeconomic inequalities, and the links between education, higher education, and the labour market.

Diego Carrasco is an Assistant Research Professor at Escuela Psicología, Pontificia Universidad Católica de Chile, where he serves as a lecturer in quantitative research methods. His research interests are methodological challenges in international large-scale assessment (ILSA) and have collaborated as a methodological expert for UNESCO-OREALC, UNESCO-UIS, UNICEF-MENA, and the OECD on ILSA studies.

Giuseppe Caruso is an economist whose research interests lie in applied microeconomics, the economics of education, labour economics, and inequality. His work focuses on educational transitions, university outcomes, social mobility, and the role of family background and institutional factors in shaping individual educational and labour-market trajectories.

Giorgia Casalone is an Associate Professor of Public Finance at the University of Piemonte Orientale, an External Researcher at the Dondena Centre for Research on Social Dynamics and Public Policy at Bocconi University, and a member of the Interuniversity Center for Regional and Local Finance (CIFREL) at the Catholic University of Milan. Her research focuses on the economics of education and the analysis of public policies.

Clelia Cascella (PhD in Research Methodology; PhD in Economics) is a senior researcher at INVALSI. Her research focuses on large-scale standardized assessments and educational inequalities, with particular attention to socio-economic measurement. Qualified as Associate Professor in Research Methodology and Mathematics, she has published widely in international journals and contributed to several European research projects.

Lorenzo Cerullo Bruno holds an M.Sc. in Economics, Management and Institutions from the University of Eastern Piedmont (109/110), with a Double Degree in Public Economics and Public Finance from the University of Fribourg. He recently completed a traineeship at the European Commission's DG EAC, working on impact evaluation of EU-funded interventions with quasi-experimental methods and large-scale datasets.

Augusto Cerqua is Associate Professor at the Department of Social Sciences and Economics, Sapienza University of Rome. His research interests include regional and industrial policy, education and labour economics, political economy, and policy evaluation. His work applies counterfactual and data-intensive methods to assess public policies and their effects across territories, firms, workers, and students.

Tiziana Chiappelli (PhD) is researcher and professor at the University of Florence, Department of Sciences of Education and Psychology. She is specialized on education, inclusive and intercultural education, minorities (including gender and migration issues), design of innovative curricula and training paths.

Francesco Chiaradia is a PhD candidate in the Department of Business Economics at Roma Tre University. His research examines socio-economic and educational inequalities through spatial and quantitative analysis at multiple territorial scales. His current work focuses on educational poverty, accessibility to urban services and the integration of official statistics and geospatial data to support the policy maker.

Emilia Ciccia is a primary school teacher at the *I.C. Praia a Mare*, where she also holds strategic roles including *Funzione Strumentale* (Institutional Support Function), INVALSI Coordinator, and member of the digital innovation team (*animatori digitali*). She holds a Ph.D. in Training Models and is a Developmental Psychologist. She integrates her fieldwork with scientific research and university-level teaching in Special Education.

Jose G. Clavel is Profesor Titular de Universidad in the Area of Quantitative Methods at the University of Murcia (Spain). Doctor since 1997 with a thesis on the application of Correspondence Analysis and

Classification and Regression Trees to bank's customer segmentation, he has been visiting professor at The Ontario Institute for Studies in Education (Toronto, Canada), The Universitat Pompeu Fabra (Barcelona, Spain) and the Indira Gandhi Institute of Development Research (Mumbai, India). Currently he is also Adjunt Professor at Kemmy Business School (University of Limerick, Ireland).

Adam Coates is a PhD student at the University of Bath. His research focuses on mathematics learning and achievement, and the use of cognitive diagnostic assessment and modelling. His research interests also include sustainability and climate change education.

Eugenia Cognigni is a second-year PhD student in Educational Research and Evaluation within the National PhD Programme in Teaching and Learning Sciences at the University of Macerata, funded by INVALSI. Her research focuses on socio-emotional factors as predictors of Implicit School Dropout. Specifically, she investigates whether these factors emerge as early indicators of academic vulnerability in Lombardy's primary schools, as reflected in teachers' representations and educational practices.

Simone Consegnati is a researcher at INDIRE, he works on school innovation, Service-Learning, and the development of transversal competencies. His studies focus on active teaching practices, student participation, educational assessment, and documentation and analysis processes in school networks engaged in innovation.

Dalit Contini is Full Professor of Social Statistics at the Department of Economics and Statistics "Cognetti de Martiis" at the University of Torino. Her research focuses on educational inequalities, education policies, and counterfactual analysis. She has extensive experience working with both survey and administrative data sources at national and international levels.

A***ntonella Coppi*** is Associate Professor of General and Social Pedagogy and Dean of the Faculty of Education of the IUL Online University. She deals with Lifelong Education and lifelong learning, inclusion and care strategies in different ages and contexts, community development, creative spaces for learning, training, Artistry, interconnections between artistic languages and intergenerational relationships in educational contexts.

Annalisa Cristini is Full Professor of Economics at the Department of Economics of the University of Bergamo and delegate of the Rector for Welfare and Sustainable Development. She attained the D.Phil. in Economics in 1990 at the University of Oxford where she was a member of Lincoln College and of the Institute of Economics and Statistics.

Clementina Crocè is a postdoctoral researcher in the Department of Economics and Management of the University of Padua. She is an applied economist, and her research focuses on education, development, gender and labour economics.

Chiara Rita Cunsolo is a PhD student in Statistical Sciences at the Department of Statistical Sciences at the University of Bologna. Her research focuses on latent variable and mixture item response theory models for educational assessment, with particular interest in test-taking engagement, disengagement, and low-effort response behaviour in cognitive and non-cognitive data.

Anna Cristina D'Addio serves as the chief of education policy for the Global Education Monitoring Report (GEM Report) at UNESCO where she also leads the PEER profiles project, and the gender report. Prior to this position, Anna worked at the OECD for over a decade on a wide range of issues with a particular emphasis on the role of policies. Earlier in her career Anna was a research professor in micro-econometrics applied to labour market/education issues. She has conducted research in several universities and research centers including the Center for Applied Econometrics (CAM), University of Copenhagen; the Higher Institute for Labour Studies (HIVA), Catholic University of Leuven; the Department of Economics, University of Aarhus; the Center of Operations Research and Econometrics (CORE), and IRES of the Catholic University of Louvain-la-Neuve. Economist by background, she holds a PhD in Quantitative Economics (European Doctoral Program) from CORE and IRES at the Catholic University of Louvain-la-Neuve (Belgium), and a PhD in Public Economics from the University of Pavia (Italy).

Martina D'Amato holds a Master in Social Data Science from the University of Milano-Bicocca.

Cristina Davino is a Full Professor in Statistics at the University of Naples Federico II, Italy. She is the coordinator of the project "Barriere invisibili: Progetto per la misurazione della povertà educativa a Napoli e provincia" in cooperation with Save the Children and of the project "Database on university students", both financed by the Next Generation EU program (2022-2025).

Simone Dei Pieri is a PhD candidate in Economics, Management and Decision Making at the University of Catania and holds a Master in Social Data Science. His research covers artificial intelligence, political and institutional communication, and the digital transformation of cultural systems.

Sonia Della Monaca holds a master's degree in Sociology and a postgraduate program in Data Analysis. She is currently a research fellow at the Department of Economics and Statistics of the University of Turin, where she is working on the career trajectories of Italian workers.

Marco Delprato is a visiting research fellow in the Department of Education at the University of Bath (UK). His research focuses on the application of quantitative analysis within the education and international development (EID) field, highlighting diverse inequalities in the global south's education systems.

Marta Desimoni is a Senior Researcher at INVALSI, Scientific Coordinator of the Methodology and Psychometrics Unit for the National Assessments, and an elected member of the Scientific Council. She coordinates the psychometric aspects of digital item banking, standard-setting procedures for English assessments, proficiency-level determination in Italian and Mathematics, and the linking of National Assessments across time to support trend analyses.

Lorenzo Di Biagio, Msc and PhD in Mathematics, is a researcher at the Italian National Institute of Statistics (Istat). At Istat, he coordinates the design of integrated well-being and sustainability analyses, co-authors annual well-being and SDG reports and manages SDG Goals 1 and 10. His current research focuses on territorial disparities, monitoring SDG performance and Machine learning-based SDG classification.

Melisa Diaz Lema is an Assistant Professor at Politecnico di Milano. Her work explores the intersection of education economics and digital innovation. Her research focuses on leveraging data to support decision-making in the public sector. She is engaged in projects that apply data-driven methodologies to improve public services, especially in education.

Theresa Dicke's research examines how personal resources, particularly self-beliefs, relate to occupational well-being, stress development, and health maintenance. Drawing on her background in instructional psychology, she also investigates learning-related constructs such as professional knowledge, student attrition, self-regulated learning strategies, and cognitive load. In her professorial role at IPPE, her work focuses especially on self-concept and related constructs in education.

Adriana Di Liberto is professor of Economics at the University of Cagliari. She holds a Ph.D. in Economics from UCL (London) and the University of Siena. She is a Research Fellow at CRENoS, RFBerlin and IZA@LISER. Her research covers the economics of education, policy evaluation and growth, with works published in leading economics and education journals.

Pasquale di Padova is a researcher at Istat within the Central Directorate for Social Statistics and Welfare, Integrated Labour, Education and Training Service. He holds a PhD in Methodology of Social Sciences. His research addresses occupational dynamics, youth labour market conditions, educational inequalities, and the integration of foreign-born populations. He is co-coordinator of the research project from which this contribution originates.

Maria Laura Di Tommaso is Full Professor of Economics at the Department of Economics and Statistics "Cognetti de Martiis", University of Turin. She is a Fellow of the Collegio Carlo Alberto and an Associate Researcher at the Frisch Centre for Economic Research, Oslo. Previously, she was a College Lecturer in Economics at Robinson College and a Research Associate in the Department of Applied Economics at the University of Cambridge. Her research focuses on gender and feminist economics. Her current research projects examine gender differences in education, particularly in mathematics and science in Italy (MathGenderGap project), the effects of the COVID-19 pandemic on educational inequalities, and the evaluation of financial incentives for the integration of migrants in Norway.

Vincent Donche is a full professor at the Department of Training and Education Sciences at the University of Antwerp, Belgium. He conducts research in the domains of higher education, learning and instruction, and educational measurement. In particular, he researches individual differences in learning (such as self-efficacy, motivation, and self-regulation). For his publication record, see: https://orcid.org/0000-0002-9405-3896.

Silvia Duranti is a researcher at IRPET, the Regional Institute for the economic planning of Tuscany (Istituto Regionale per la Programmazione Regionale della Toscana). Her work focuses on the economics of education and training, with particular attention to the territorial analysis and to the evaluation of public policies.

Camila Echeverría is a Psychometrician at DEMRE, Universidad de Chile, the agency responsible for Chile's national university admissions test (PAES). Her work focuses on operational item calibration, equating, and quality control for large-scale, high-stakes assessment, including the development of procedures for detecting and handling aberrant response behaviour in operational calibration samples.

Ivan Etzo is Associate Professor of Applied Economics at the Department of Economics and Business, University of Cagliari, and a research affiliate at CRENoS. His research focuses on regional economics, migration, tourism, and territorial development. He studies mobility and regional disparities using applied microeconometric and spatial approaches, with particular attention to the Italian context.

Nurullah Eryilmaz is a Senior Research Analyst at the International Association for the Evaluation of Educational Achievement (IEA), Hamburg, and a postdoctoral researcher at Leuphana University. His research focuses on international large-scale assessments, educational inequality, longitudinal analyses, educational leadership, and the application of advanced quantitative methods and machine learning to comparative education.

Pablo Espinoza is a Psychometrician at DEMRE, Universidad de Chile. His work centres on the operational scaling and calibration of the PAES admissions battery, item banking, and the implementation of psychometric quality-control procedures under operational constraints, with particular interest in the effects of guessing and disengaged responding on item parameter estimation.

Laura Evangelista is a researcher at the National Institute for the Analysis of Public Policies (INAPP). Her work focuses on vocational education and training, quality assurance, Peer Review, evaluation practices and the development of tools supporting improvement processes in education and training systems.

Chiara Fante is a Researcher at the Institute for Educational Technology of the Italian National Research Council (CNR-ITD), Genoa. A psychologist with a PhD in Clinical Psychology, her research spans educational neuroscience, school inclusion, and technology-supported learning. She is the SCHOOLWISE Scientific Coordinator for pilot and evaluation activities.

Emanuele Fedeli is a postdoctoral researcher at the Department of Social and Political Sciences at the University of Milan, where he works on the Horizon Europe LEARN project. He holds a PhD in Sociology and Social Research from the University of Trento and has also been a VisitINPS fellow, working with Italian administrative data on childcare, labour market inequalities, and family policies. His research lies at the intersection of the sociology of education, social stratification, and policy evaluation, with a particular focus on peer effects, educational inequalities, grading standards, and territorial disparities in educational

opportunities. His work has been published in journals including Journal of Ethnic and Migration Studies, Social Science Research, European Societies, Journal of International Migration and Integration, Social Indicators Research, and Quality & Quantity.

Luigi Fenza is a doctoral candidate in Human Sciences at the University of Perugia and a research fellow in Life Plan, Adult Disability and Inclusive Education. His research interests include inclusive and special education, gifted education, and the design and evaluation of educational interventions supporting the participation and development of learners with diverse educational needs.

Diana Fernández Méndez, PhD, Assistant Professor in the Department of Fundamentals of Economic Analysis and researcher in GAME-IDEGA at Universidade de Santiago de Compostela (Spain). Her research focuses on labour economics, public policies, social services, tourism and regional socioeconomic analysis. She has published on the labour market effects of long-term care policies, population ageing, and the socioeconomic impact of the Camino de Santiago.

Noemi Ferrara is a junior researcher at IRES Piemonte. She holds a master's degree in Statistics from the University of Turin and works on higher education, student support policies and university-to-work transitions.

Fabio Ferri holds a Master in Social Data Science from the University of Milano-Bicocca.

Letizia Ferri is a RTT, researcher in Tenure Track at the University of Insubria (Italy). Her work spans educational research, teacher professional development, instructional design and the design of training pathways. Her interests include Evidence-Based Education, teacher education, and the use of artificial intelligence for qualitative data analysis and learning assessment.

Valentina Ferri holds a PhD in Population and Development Economics and a Master's degree in Public Policy Analysis and Evaluation. Since 2017, she has been a Researcher at INAPP, leading the research group on Technologies and Artificial Intelligence. Her work focuses on skills, labor markets, education-to-work transitions, and digital and green transformations. She has coordinated major research projects and authored more than 70 scientific publications.

Giovanna Filosa, social psychologist and psychotherapist of the developmental age with a psychodynamic focus, PhD in Psychology of communication, socialization and interaction, as first technologist has carried out various research activities on educational policies, migration policies and educational poverty. She currently works at the research group "Technologies and Artificial Intelligence: work, professions, training and skills" of INAPP.

Sebastiana Fisicaro is a PhD candidate in Education at the University of Córdoba (Spain) and a former technical officer at the Italian Ministry of Education. Her research interests include school self-evaluation, data use, school improvement, and educational innovation. She has contributed to the creation of school networks and is involved in professional development through the cultural association Rete Sophia 3.0.

Concetta Fonzo is a researcher at the National Institute for the Analysis of Public Policies (INAPP). Her research interests include vocational education and training, quality assurance, Peer Review, evidence-informed evaluation, organisational learning and the use of data to support improvement in education and training institutions.

Daniel Frischemeier is Professor of Mathematics Education at the University of Münster, Germany. His research focuses on statistics education, data science education, statistical thinking, and technology-enhanced learning. He serves in leadership roles within the International Association for Statistical Education (IASE) and has coordinated several national and European research projects, including DataSETUP and PINGU.

Giulia Gambardella is a PhD Student in Economics at the joint doctoral program of the University of Pavia and University of Milano. She previously collaborated as a research assistant at the University of Perugia in

the TEMPO project. Her research interests are mainly related to socio-emotional skills development and early childhood interventions.

Letizia Gambi is an affiliate researcher at the Leuven Economics of Education Research (LEER) group at KU Leuven and an external expert for DG EAC. Her research focuses on teachers, student achievement, school organisation, and the evaluation of public policies using counterfactual methods. Her work has been published in a range of outlets, including book chapters, policy reports for the European Commission and the Irish government, and academic journals such as Labour Economics.

Inés García-Bohórquez is a PhD candidate at the University of Salamanca, affiliated with the EMC3 research group in the Department of Mathematics and Science Education. Her research focuses on climate change education through theoretical and statistical modelling using ILSA data. She holds a degree in Primary Education and a master's in ICT in Education.

Francisco Gatica is a Research Associate at the University of Bath, contributing to the International Civic and Citizenship Education Study (ICCS) 2027. Previously, he coordinated the Regional Comparative and Explanatory Study (ERCE) at UNESCO Santiago. He is a PhD Candidate in Education and holds a Master's in Public Policy and a Master's in Educational Research, both from the University of Chile.

Orazio Giancola is an Associate Professor at the Department of Social and Economic Sciences, Sapienza University of Rome. His research examines social and educational inequalities, social mobility, and comparative educational systems. Over two decades, he has developed expertise in international datasets (OECD: PISA, PIAAC), employed in over thirty publications. His work integrates quantitative and comparative approaches to investigate educational transformations and social stratification.

Marco Giganti holds a PhD in Educational Sciences and is currently a tenure-track Assistant Professor in Experimental Pedagogy at the Department of Human and Social Sciences, University of Bergamo. His research focuses on empirical research methodology in education, assessment theory and practice (particularly formative, standardized and through GenAI assessment), digital competences and their assessment, and teacher training (pre-service and in-service).

Ludovica Giua is Assistant Professor of Econometrics at the University of Cagliari and a HEDG and CRENoS affiliate, with expertise in applied microeconometrics and policy evaluation. Her research focuses on the economics of risky behaviours, immigration and education. She holds a Ph.D. in Economics from the University of Essex and was previously Economic Analyst at the Joint Research Centre of the European Commission.

Pamela Giustinelli is Associate Professor of Economics in the Department of Economics and Management of the University of Padua, and Associate Editor of the Economic Journal. She is an applied microeconometrician, and her research explores the microeconomics and econometrics of individuals' subjective expectations, with applications in the economics of education, the family, health, labour, climate change and firm behaviour.

Tommaso Giorgi is a PhD student in Statistics at the University of Bologna. His research focuses on item response theory models and supervised learning, with particular interest in classification and regression trees applied to item analysis, digital skills, and educational assessment processes.

Olga Gorodetskaya is researcher at FBK-IRVAPP. She holds a PhD in Sociology and Social Research. Her research focuses on labour market inequalities, work–family dynamics, and policy evaluation. She works primarily with quantitative methods and survey data.

Serena Greco holds a PhD in Learning Sciences and Digital Technologies and is a researcher at INDIRE, where she has worked since 2010. Her research focuses on school governance, leadership, middle management and innovation processes. She also teaches at IUL University and directs master's and postgraduate courses on school governance.

Elena Grimaccia is Senior Research Fellow at Istat, Rome. She holds a PhD in Economics and a Master in Social Data Science. She is a Chartered Statistician of the Royal Statistical Society, an EU COST Expert, and Associate Editor of the Journal of the Royal Statistical Society (Series A). Her research has been published across leading international journals such as Social Indicators Research, Quality and Quantity, the Journal of Official Statistics, Environmetrics, Statistical Methods and Applications.

Paolo Guandalini is a PhD student at the Department of Cognitive Sciences of the University of Trento. His research work focuses on teacher careers, specifically the sustainability of teachers' careers in the later stages of working life.

Juliana Hartz Rodrigues is a trainee at Istat who is currently pursuing a Master's degree in Statistical Methods and Applications at Sapienza University of Rome. Her professional experience combines public policy, education, and statistical analysis, with previous work at the Office for the Measurement of Learning Quality of the Peruvian Ministry of Education as well as in the Peruvian Ministry of Labour. These experiences have shaped her interest in issues related to the transition from education to work.

Mobarak Hossain is an Assistant Professor in the Department of Social Policy at the London School of Economics and Political Science. His research examines social and educational inequalities, intergenerational mobility, and education policy in comparative and global perspectives, with particular focus on low- and middle-income countries and the use of large-scale administrative and survey data.

Gabriele Iannotta is a Postdoctoral Researcher at Politecnico di Milano, Department of Management, Economics and Industrial Engineering. He holds a Ph.D. in Economics from the Catholic University of Milan. His work focuses on behavioral and experimental economics, with a particular interest in how education shapes preferences, attitudes, and decision-making. He currently conducts research on socio-emotional learning (SEL) and financial education, using randomized evaluations and experimental methods to study how schools can foster skills such as self-regulation, patience, and financial competence. His projects involve collaborations with institutions such as the Bank of Italy and the European Commission.

Ioseba Iraurgi holds a PhD in Psychology. He is a Clinical Psychologist, researcher and full Professor at the Faculty of Health Sciences at the University of Deusto, where he leads the 'Evaluation, Clinical and Health' research group and directs DeustoPsych. With 30+ research projects and 250+ publications, his expertise spans psychotherapy outcomes, psychological assessment tools, meta-analysis, and gifted education.

Maria–Magdalena Isac is the International Study Co-Director for the International Civic and Citizenship Education Study (ICCS) and a research officer in the Department of Education at the University of Bath (UK). Her work focuses is the evaluation of school systems in Europe, specifically analysing how teaching practices shape youth citizenship learning and democratic.

Miryam Martínez Izaguirre holds a PhD in Pedagogy and is an expert in gifted education. She is currently a lecturer and researcher at the Faculty of Education and Sport of the University of Deusto. As a member of eDucaR (Deusto Education Research) research team, her research focuses on educational support for gifted and talented students, teaching innovation and competency development.

Jason Jabbari is an Assistant Professor at Washington University in St. Louis, where he directs the Clark-Fox Policy Institute and the Center for Education Research, Practice, and Policy Partnerships. His research examines how policies, programs, and practices shape human capital development, social mobility, and more equitable academic and economic trajectories for marginalized populations.

Stefan Johansson is Associate Professor at the University of Gothenburg. His research focuses on educational measurement, psychometrics, teacher quality, educational effectiveness, and international large-scale assessments. He has extensive experience analysing longitudinal and comparative educational data to inform policy and practice.

Heli Kauppinen is a project researcher at the Finnish Institute for Educational Research, University of Jyväskylä, Finland. She works as a statistician in a research group on assessment of education, focusing on

international large-scale assessment studies. Her research interests include educational equity and test-taking motivation. She is currently pursuing a PhD in statistics education at the University of Jyväskylä.

Ekaterina Kochergina is a postdoctoral researcher at the University of Milano-Bicocca. Her research focuses on ageing, social inequalities, digital inclusion, and comparative life-course research. She specialises in quantitative analysis of large-scale longitudinal datasets and has contributed to several international research projects investigating social participation and well-being in later life.

Giovanni Lari is a quantitative economist currently working at the European Commission (DG EAC), where he contributes to the impact evaluation of EU programmes. He holds a Master's degree in Economics from the University of Piemonte Orientale, specializing in causal inference and Propensity Score Matching. He collaborates with Andrea Saltelli on Global Sensitivity Analysis applied to educational research.

Federica Laudisa holds a sociology degree and a master's in public policy analysis. She is a researcher at IRES Piemonte. Her main fields of study are student support policies, both at the regional and national levels. From 2017 to 2019, she was a member as an expert at the National Observatory for University Student Affairs.

Keti Lelo is Associate Professor of Economic History in the Department of Business Economics at Rome Tre University. She holds a PhD in Geomatics and in Economics and Quantitative Methods. Her research covers urban history and urban and regional economics, with a specific focus on the economic landscape and socio-spatial inequalities. She teaches History of Cities and Territories, Urban and Regional Analysis and Ethical Business Management.

Xiao Li is a research fellow at the University of Brescia, working on the SPEED project on predicting university student dropout. He holds a PhD in Economics from the Universities of Milan and Pavia, completed in collaboration with Fondazione Eni Enrico Mattei (FEEM). His research interests include energy and commodity markets, applied econometrics, and forecasting, including energy transition metals.

Francesca Licari is a Ph.D. demographic researcher at the Italian National Institute of Statistics (Istat). Her research focuses on internal and international migration, as well as statistical methods for the allocation (imputation) of missing data.

Valentina Liguori is a PhD candidate in Public Management at the Department of Management, University of Bologna, supervised by Professor Angelo Paletta. Her research focuses on Professional Learning Communities as organisational enabling conditions for value creation in public educational institutions.

Ilaria Lievore is senior program officer and social researcher at Fondazione per la Scuola (Compagnia di San Paolo) in Turin. She holds a Ph.D. in Sociology and Social Research at the University of Trento. Her research interests concern social stratification and educational inequality. Her articles have been published in international journals such as Social Forces, Social Science Research, European Societies.

Sabrina Lipari is a Program Manager and Quality Specialist at Junior Achievement Italy, with extensive experience in designing and evaluating educational programs that strengthen entrepreneurship and life skills among young people. She specializes in impact measurement and ecosystem-based approaches to develop scalable models of experiential learning aligned with European priorities on skills and employability.

Natalia López–Hornickel is a Research Associate at the International Study Centre for ICCS, University of Bath, where she earned her PhD and MRes in Advanced Quantitative Methods (SWDTP/ESRC). She also holds a Master's in Sociology from the Pontificia Universidad Católica de Chile. Her work examines attitudes towards gender equity using large-scale data and advanced statistical methods across Latin America, Europe, and the Middle East.

Andrea Lopes is a third-year PhD Student in Economics at the University of Milan. He has been a visiting researcher at Bocconi University and at the ifo Center for the Economics of Education and to the CESifo

network. His research focuses on the Economics of Education. His research has been recognized with the Marco Manera Research Prize and a grant.

Samuele Longo is a Master's student in Statistics and Data Science at the University of Palermo. He focuses on data analysis and statistical modeling, with interests in inference, machine learning, and applications of statistics in economic and social contexts. He has skills in statistical programming and data analysis.

Gabriele Lombardi is a Research Fellow in Social Statistics at the University of Florence. His research interests include internal migration of Italian university students, university-to-work transition, adolescent gambling addiction, and the propensity to adopt pro-sustainable behaviours among adolescents.

Patrizia Lotti is a researcher at INDIRE, she works on Service-Learning, didactic innovation, and teacher education. Her research and support activities focus on school practices oriented toward participation, active citizenship, and the relationship between school, territory, and teachers' professional development.

Mariavittoria Ludovico is a research fellow at Politecnico di Milano, focusing on the economics and management of education and on student outcomes. Her main strand of research concerns early school leaving and low performance, which she studies using quantitative methods on large-scale administrative and standardised-test data. She also works on the right to education, inclusion, and digital transformation.

Barbara Luppi is Associate Professor of Economics at the Marco Biagi Department of Economics, University of Modena and Reggio Emilia. Her research interests include behavioural economics, law and economics, decision-making under uncertainty, and the economic analysis of institutions. Her work combines theoretical and applied approaches to investigate individual behaviour and institutional outcomes.

Marta Magnani is a PhD student at the University of Milan and is affiliated with the Milan Higher Education Observatory (MHEO). Her research focuses on distance-learning students' trajectories and school-to-university transition. She has also gained research experience at the OECD Centre for Skills and the Italian Ministry of University and Research.

Maria Rita Mancaniello, associate professor of Social Pedagogy, spent her academic career at the University of Florence in the FORLILPSI department until 2023, and began working at the University of Siena in the DISPOC department in 2023. Her theoretical and experimental research spans distinct fields, yet they are linked by a common thread rooted in their relational nature.

Flavio Manganello is a Senior Researcher at the Institute for Educational Technology of the Italian National Research Council (CNR-ITD), Genoa, with a PhD in Engineering Sciences (e-learning). His research addresses AI governance in education, accountability under the AI Act, learning analytics, and self-regulated learning. He is Principal Investigator and coordinator of the SCHOOLWISE project.

Lucia Mangiavacchi is Associate Professor of Economics at the University of Perugia and IZA Research Fellow. Her current research explores the economics of parenting and child development. It examines the impact of parental decision-making on children's human capital investment and the role of children's and adolescents' time allocation in the production of cognitive and non-cognitive skills.

C*ostanza Marconi* is a Post-Doc Researcher at the Department of Economics of the University of Bergamo. She holds a Ph.D. in Institutions and Policy from the Catholic University of Milan (2023). Her research interests are in public economics and political economy, with a particular focus on public finance, education, and gender-related issues.

Lorenzo Maraviglia holds a PhD in Development Sociology from the University of Pisa. From 2010 to 2022, he served as Head of the Statistical Office of the Province of Lucca. Since 2022, he has been working as a Statistical Researcher at INVALSI, contributing to the production of national reports on student achievement and competencies.

Herbert W. Marsh is ACU MacKillop Distinguished Professor and one of the world's most highly cited educational researchers, with 800+ publications and over 209,600 citations. His influential work spans self-

concept, motivation, educational effectiveness, bullying, multilevel modelling, and structural equation modelling. A decorated fellow of major scholarly societies, he founded the international SELF Research Centre and advanced substantive-methodological research synergy globally.

Francesco Marsili is post-doc researcher at the University of Perugia, Department of Philosophy, Social Sciences, Humanities and Education. His research focuses on gifted education, special and inclusive education, and research synthesis methods, including systematic reviews and meta-analyses.

Michele Marsili graduated in Statistical, Demographic and Economical Sciences at the University of Rome "La Sapienza". Since 2018, he has been working in the Statistical Department at INVALSI. His current research interests focus on the application of large language models (LLMs) to educational assessment.

Sonia Marzadro is researcher at FBK-IRVAPP. Her research focuses on social inequality, including income distribution, poverty, social mobility, and educational opportunities, policy evaluation. She holds a PhD in Sociology and Social Research.

Chiara Masci is a tenure-track researcher in Statistics at the Department of Economics, Management and Quantitative Methods, University of Milan, and is affiliated with the Milan Higher Education Observatory (MHEO). Her research focuses on statistical modelling for longitudinal and hierarchical data, with applications to education, student dropout, health data and risk prediction.

Giancarlo Masi is a Research Associate at the Institute for Educational Technology of the Italian National Research Council (CNR-ITD), Genoa, and a PhD candidate in Learning Sciences and Digital Technologies (University of Modena and Reggio Emilia), working on AI-enhanced adaptive learning. He is the SCHOOLWISE Quality Assurance Manager.

Jorge Matamala is a research assistant at Institute of Socio-Emotional Well-being (IBEM) Universidad del Desarrollo, Santiago, Chile. His research focuses on citizenship education and teacher education, with a specific emphasis on the development of citizenship competences among student teachers.

Eleonora Mattarelli is a Research Assistant at Istat within the Central Directorate for Social Statistics and Welfare, Integrated Labour, Education and Training Service. She holds a PhD in Educational Research from the Department of Developmental and Socialization Psychology Processes, Sapienza University of Rome. Her expertise in educational research derives from both her doctoral work in experimental pedagogy and her experience at INVALSI.

Mariagiulia Matteucci is an Associate Professor of Statistics in the Department of Statistical Sciences at the University of Bologna. Her main research interests include item response theory models, Bayesian estimation methods, and statistical applications in educational measurement and evaluation.

Maria Luisa Maitino is a researcher at IRPET, the Regional Institute for the economic planning of Tuscany (Istituto Regionale per la Programmazione Regionale della Toscana). Her work focuses on poverty and income distribution, using microsimulation models and counterfactual evaluation methods to assess public policies in these fields.

José Manuel Amoedo Meijide, PhD, Assistant Professor in the Department of Fundamentals of Economic Analysis and researcher in GAME-IDEGA at Universidade de Santiago de Compostela (Spain). His research focuses on income inequality, educational inequalities, digital vulnerability, and regional development. He has published in several international peer-reviewed journals and collaborates with researchers across Europe.

Maria Meletiou-Mavrotheris is Professor of Mathematics Education at European University Cyprus and Director of the ICT-Enhanced Education Laboratory. Her research focuses on mathematics and statistics education, educational technology, STEM education, and educational data science. She has coordinated numerous internationally funded projects, including award-winning game-based and digital learning initiatives recognized with the International Association for Statistical Education Best Cooperative Project Award and two European Innovative Teaching Awards.

Andrea Meucci holds a PhD in Physics and is a Mathematics and Science teacher at I.C. "Dante Alighieri" in Aulla. He serves as the official institutional Coordinator for the INVALSI National Assessment Evaluation.

Alessandra Michelangeli is Associate Professor of Economics at the University of Milan-Bicocca. She has extensive experience in inequality studies She has published several academic articles in international economic journals and has served as a consultant for the European Commission and other Institutions. She has been recently involved in a PRIN 2022 PNRR project on health inequalities and two Cariplo Inequalities Research Projects. ORCID ID: https://orcid.org/0000-0001-6802-6870.

Stefania Mignani is a Full Professor of Statistics in the Department of Statistical Sciences at the University of Bologna. Her primary research interests focus on latent variable models and statistical methods for social and educational phenomena, with particular attention to the assessment of competencies.

Maria Mirandi is a researcher at the Sophia University Institute. Her research interests include the adaptation and validation of psychometric instruments in educational contexts.

Daniel Miranda is an assistant professor of the Department of Psychology at the Universidad de Chile, Associate research at the Center for Social Conflicts and Cohesion Studies - COES and adjunct researcher at the Millennium Nucleus of Opportunities and Digital Inequalities - NUDOS, Chile. His research focuses on citizenship education, political inequalities and their intergenerational transmission.

Giulia Montefiore is a PhD student at the Department of Sociology at the Autonomous University of Barcelona, Spain. Her doctoral work centers on the political sociology and school enactment of school autonomy and accountability policies in Italy, with a specific focus on data use by schools and teachers. Giulia's research interests also include educational inequalities and school segregation.

Daniele Montepara holds a Master in Social Data Science from the University of Milano-Bicocca and is a Full Stack Developer at Capgemini, with over twenty years of experience in web application development across front-end and back-end technologies, multiple programming languages, frameworks and cloud platforms.

Iacopo Moreschini is a PhD candidate in the Department of Sociology and Social Research at the University of Milano-Bicocca and holds a Master in Social Data Science. His research addresses educational inequalities, teacher practices and student achievement, drawing on INVALSI data, with recent publications in Quality & Quantity and Scuola Democratica.

Annalisa Morganti is an associate professor at the University of Perugia, Department of Philosophy, Social Sciences, Humanities and Education. Her research focuses on inclusive education, social-emotional learning, and evidence-based educational interventions. She has coordinated national and international research projects on inclusion and collaborates with international networks on social-emotional competence development in schools.

Sara Mori holds a PhD in Evaluation of Educational Systems and Processes and is a psychologist-psychotherapist. She is a researcher at INDIRE, where her work focuses on transversal competences, evaluation of educational processes, school well-being and inclusion. She teaches at IUL University and has worked on national and international projects on personalised learning, educational neuroscience and educational technologies.

Léonard Moulin is a tenured researcher in economics at the Institut national d'études démographiques (INED). He completed his PhD in 2014 on the effects of tuition fees in French higher education and obtained his habilitation from Université Paris-Est Créteil in 2024. His research addresses the economics of education and educational inequalities, from school to adulthood. He is co-investigator of the Horizon Europe LINEup project on longitudinal data for educational inequalities, and previously held the Île-de-France COVID-19 and Education Research Chair.

Jurgena Myftiu is an Assistant Professor of Economic Statistics at the Department of Economics, University of Bergamo. Her research focuses on socio-economic inequalities, education, labour markets, tourism,

mobility and accessibility, with a strong emphasis on applied quantitative methods and the construction of composite indicators. She has participated in several competitive research projects and has extensive experience in the analysis and harmonisation of large-scale microdata.

Chiara Mussida is an Associate Professor of Economic Policy at Università Cattolica del Sacro Cuore (Piacenza, Italy) and a Research Fellow at the Global Labor Organization (GLO). Her research interests include empirical labour economics, with a particular focus on unemployment, gender inequality, poverty, disability, and other forms of discrimination.

Marcello Napoli graduated in Marketing Analytics and Metrics at the LUISS Guido Carli University in Rome, has a 1st level Master's degree in Digital Strategy & Data Analytics at the 24Ore Business School. He worked as a consultant in CEM and CRM fields. Since February 2024 he has been working at INVALSI in the INVALSI Statistical Department.

Enrico Nerli Ballati is a researcher at Istat within the Central Directorate for Data Collection, Data Collection for Surveys Without Interviewing Networks Service. He holds a PhD in Methodology of Social Sciences from Sapienza, University of Rome. Over the years, he has dealt with youth conditions in the labor market and has developed expertise in the thematic area of educational research during his work experience at INVALSI, in the Innovation and Development and International Surveys areas.

Trude Nilsen is Professor of Educational Measurement at the University of Oslo. Her research focuses on educational effectiveness, teaching quality, school climate, science and mathematics education, and international large-scale assessments, particularly TIMSS and PISA. She has led numerous international research projects on educational quality and equity.

Fiammetta Noccioli graduated in Statistical, Demographic and Economical Sciences at the University of Rome "La Sapienza". She obtained the II Level Master's degree in big data and machine learning. She previously worked as a statistician in epidemiological areas. Since November 2023 she has been working in the INVALSI Statistical Department.

Federica Origo is Full Professor of Economics at the Department of Economics of the University of Bergamo (Italy). She received her Ph.D. in Economics from the Catholic University of Milan (2004), and her M.Sc. in Economics from the University of Texas at Austin (2001). Her research interests are in the fields of labour, economics of education, and treatment evaluation.

Lorenza Orlandini, PhD, is a researcher at INDIRE. Her main research interests include educational approaches that promote the idea of an open school connected to its local area, flexible models for organizing school time, and innovative, inclusive learning environments*.*

Francesco Ottaviani is a Statistical Officer for Public Policy Evaluation at Ministry of Economy and Finance - State General Accounting Department. His research interests include Programme and Policy Evaluation, Large-scale Assessment and Labour Market.

Angelo Paletta is Full Professor of Business Administration at the Department of Management, University of Bologna, where he serves as Department Director. His research focuses on school leadership, Professional Learning Communities, and value creation in public educational institutions. He is President of IPRASE (Trento) and co-director of the Master's in Public Management and Innovation at Bologna Business School.

Laura Palmerio is a Senior Researcher at INVALSI, where she leads the International Surveys Area and coordinates Italy's participation in IEA TIMSS, PIRLS, and ICCS, and OECD PISA and TALIS. She holds a PhD in Innovation and Evaluation of Educational Systems. Her research focuses on equity, reading and mathematics competencies, civic education, citizenship, and information literacy.

Valentina Patacchini is researcher at IRPET**,** the Regional Institute for the economic planning of Tuscany (Istituto Regionale per la Programmazione Regionale della Toscana. Her work focuses on firms and labour statistics, based both on survey and administrative data. She has contributed to several IRPET studies on education and training pathways, labour-market entry and graduates' employment outcomes in Tuscany.

Serafina Pastore, Ph.D. in Instructional Design and Educational Assessment and Fulbright Research Fellow, is an Associate Professor at the University of Bari in Italy. Her research interests include Formative Assessment, Feedback Practice, Self-assessment and Teacher Assessment Literacy.

Veronica Pastori PhD in Methodology of Social Sciences. Her research interests include social inequalities, migration, evaluation of educational processes, data quality, and standardised questionnaire development. She previously worked at INVALSI and currently works as a researcher at ISTAT (Italian National Institute of Statistics), Rome, Italy.

Fabrizio Patriarca is Associate Professor of Economic Policy at the Marco Biagi Department of Economics, University of Modena and Reggio Emilia. His research focuses on inequality, human capital, education, labour economics, and economic policy. He has collaborated with national and international institutions and has led research projects on inequality and human capital.

Marta Pellegrini is an Associate Professor at the University of Cagliari, Department of Pedagogy, Psychology, Philosophy. Her research interests include evidence-based education, systematic reviews and meta-analyses of educational interventions, and the use of research evidence in educational policy and practice.

Eufrasio Pérez Navío is Professor in the Department of Pedagogy at the University of Jaén, where he currently serves as Head of Department. His research focuses on teacher education, didactics, educational innovation, educational technologies, school organisation, and higher education. He has authored numerous scientific publications and has extensive experience in academic research, doctoral supervision, and international educational projects.

Daniela Piazzalunga is Assistant Professor of Economics at the Department of Economics and Management, University of Trento. She is affiliated with IZA and CHILD–Collegio Carlo Alberto. Previously, she was a postdoctoral researcher at FBK-IRVAPP after earning her PhD in Economics from the University of Turin. Her research interests lie in labor economics, gender inequalities, family economics, and policy evaluation.

Luca Piccoli is Associate Professor of Economics at the University of Trento and IZA Research Fellow. Luca's current research focuses on family economics and the economics of risky behaviors, particularly the role of parents in socio-emotional skill formation among young adults, and how endogenously varying time preferences may impact the development of risky behaviors.

Giulia Pietrangeli graduated in Computer Science from Sapienza University of Rome and is currently pursuing a Master's degree in Computer Science at the same university. Her interests include artificial intelligence, machine learning, and large language models.

Giuseppe Pignataro Associate Professor at the Department of Economics of the University of Bologna and member of the Social Impact Agenda for Italy. His research interests focus on the study of inequalities of opportunity and their measurement, as well as on the causal inference evaluation of public policies, with particular reference to the issues of education and educational processes in general.

Giovanni Piumatti is a Research Officer at Fondazione Agnelli. He obtained a PhD in Psychology from the University of Turin and has worked at the School of Public Health in Belgrade, the University of Oxford, the University of Geneva, the Geneva University Hospitals and the Università della Svizzera italiana. His main research interests include health psychology, social sciences, and education.

Samuele Poy is an Assistant Professor (RTD-B) in Economic Policy in the Department for Sustainable Development and Ecological Transition and Adjunct Professor at the University of Piemonte Orientale (UPO), Vercelli (Italy). His research mainly focuses on counterfactual impact evaluation, public policy analysis, labour and social policies, welfare.

Francesca Prever graduated with a degree in Statistical Sciences from the University of Turin, where she is currently working as a research fellow. Her current research focuses on university-to-work transitions, while her broader research interests include synthetic data generation for privacy protection.

Andrea Priulla received his PhD in Social Statistics from the University of Palermo in 2023. He is currently a Tenure-Track Researcher in Statistics at the Department of Economics and Law of the University of Enna KORE. His research focuses on socioeconomic, gender, ethnic, and geographical inequalities in Italian higher education.

Jonna Pulkkinen, PhD in Educational Sciences, works as a senior researcher at the Finnish Institute for Educational Research, University of Jyväskylä. Her research is related to educational evaluation and equality in education, and she has utilized international large-scale assessment data in her research. Pulkkinen has the title of docent (education, especially educational evaluation) at the University of Turku.

Martina Ragusa is a research fellow at the Institute for Educational Technology of the Italian National Research Council (CNR-ITD), Genoa. She holds a Master's degree in Digital Humanities from the University of Pisa, and her research focuses on digital learning design, generative AI, governance, and self-regulated learning.

Alicia Ramos is a postdoctoral researcher affiliated with the KU Leuven and the University of Antwerp in Belgium. Her research focuses on motivation and learning development, underachievement among high-ability students, and the role of personal and contextual factors in shaping students' educational trajectories, particularly in transitions to secondary and higher education. For her publication record, see: https://orcid.org/0000-0002-1701-2070.

Letizia Ravagli is a researcher at IRPET, the Regional Institute for the economic planning of Tuscany (Istituto Regionale per la Programmazione Regionale della Toscana), in the research area "Labour, Inequality and Welfare". Her main research interests include income inequalities among workers and households, poverty and social exclusion, and fiscal, welfare and labour policies.

Ilaria Ravasi, holds a PhD in Human and Educational Sciences and is currently a research fellow at the University of Bergamo. She is developing a project titled "Digital competencies and school dropout in upper secondary education: an empirical analysis and the design of guidelines to promote equity."

Catalina Reszczynski holds a degree in Psychology and a Master's in Educational Psychology, both from the Pontificia Universidad Católica de Chile. Her research focuses on cognition and creativity, employing quantitative methods to study how these processes develop and can be measured in educational settings, contributing to work applying advanced statistical modeling to large-scale psychological and educational data.

Patrizia Rinaldi is a lecturer at the University of Granada (Spain), where she received a PhD (Cum Laude) in Migration Studies. Patrizia is a fellow of the Refugee Law Initiative (London). She is an active member of three COST Actions (EU). Her internship was at the Centre for Migration Studies in New York (US). Patrizia has published internationally in leading academic journals.

Lorenzo Rocco is Professor of Economics in the Department of Economics and Management of the University of Padua, and co-Editor of Economics of Education Review and Review of Economics of the Household. He is an applied economist, and his research is mainly in education, health and labour economics.

Barbara Romano is a researcher at Fondazione Agnelli. Her work focuses on the design and evaluation of education policies, with particular expertise in counterfactual evaluation and randomized controlled trials (RCTs), as well as analysis based on national and international learning assessments. She holds a PhD in Evaluation of Educational Policy at University of Genova and an MSc from the University of Pennsylvania.

Rosaria Romano is an associate professor of Statistics at the University of Naples Federico II. Her research interests include multi-block data analysis, which involves the analysis of multiple data tables that share a

common dimension. Her scientific activity has consistently focused on the implications of methodological analysis in various research areas: sensory analysis, consumer analysis, marketing, and education.

Maximiliano Romero is a Psychometrician and Research Analyst at the International Association for the Evaluation of Educational Achievement (IEA). His work focuses on international large-scale assessments, educational measurement, sampling, and cross-national research. His current research examines attitudes and achievement on a global scale, with particular attention to mathematics self-concept, educational inequalities, and cross-country variation in student outcomes.

Camilo Ruiz is a Professor at the Science Education Department at the Salamanca University; he is the PI of the EMC3 research group which study the interaction between climate change and education. His research focuses on sustainable and climate competences, from curricular design to implementation of assessment and its relation to science education.

Giorgia Ruzzante is a fixed-term researcher at the Free University of Bozen-Bolzano. Her research focuses on inclusive pedagogy and inclusive teaching**.**

Francesco Salomone Marino holds his PhD in Economics and Statistics at the University of Palermo. He has been visiting scholar at the University of Lancaster (UK). His research is focused on social mobility, income mobility and education. He is currently a civil servant at the Ministry of University and Research.

Jackiline Brigida Salazar Torres, PhD, Assistant Professor in the Department of Quantitative Economics and researcher in GAME-IDEGA at Universidade de Santiago de Compostela (Spain). Her research focuses on territorial development and labour economics, with particular attention to human and social capital, well-being, job quality, socioeconomic inequalities, and public policy evaluation. he has extensive experience in European research projects and international collaboration initiatives with Latin American institutions.

Marilena Salsano is a School Principal with a Ph.D. focused on Educational Effectiveness. An expert in the analysis of institutional self-evaluation processes and the development of quality management systems, she combines advanced pedagogical and managerial skills to promote the continuous improvement of educational institutions, optimize resources, and enhance merit through evidence-based strategies and standards of excellence

Andrea Saltelli (www.andreasaltelli.eu) has worked on physical chemistry, environmental sciences, applied statistics, impact assessment and science for policy. His main disciplinary focus is on sensitivity analysis of model output, a discipline where statistical tools are used to interpret the output from mathematical or computational models, and on sensitivity auditing, an extension of sensitivity analysis to the entire evidence-generating process in a policy context.

Guido Salza is a researcher at the Department of Social and Political Sciences of the University of Milan. His research focuses on the sociology of education, with a particular interest in educational inequalities in school achievement. His work has appeared in international journals such as the European Sociological Review and Research in Social Stratification and Mobility, as well as in the Italian journal Scuola Democratica.

Silvia Salini is Full Professor of Statistics at the Department of Economics, Management and Quantitative Methods, University of Milan, and is affiliated with the Milan Higher Education Observatory (MHEO). Her research covers robust statistics and statistical learning, with applications to education, labour-market outcomes, data science and policy-relevant empirical research.

Andres Sandoval Hernandez is Professor of Policy and Educational Assessment at the University of Bath and International Study Director of ICCS. His research focuses on educational policy, international comparative assessment, civic and citizenship education, and educational inequality. He previously led the IEA's Research and Analysis Unit and has collaborated with OECD, UNESCO, UNICEF, and the World Bank.

Giulia Savio is Assistant Professor of Economics at the Department of Economics and Statistics "Cognetti de Martiis", University of Turin, and an Affiliate of the Collegio Carlo Alberto. She is also an External Fellow of the AXA Gender Lab at Bocconi University, where she previously worked as a postdoctoral researcher. During her PhD in Economics at the Università della Svizzera italiana, she was a visiting researcher at the European Commission's Joint Research Centre and at Universitat Pompeu Fabra. Her research interests lie in public economics, gender economics, economics of education, and political economy.

Miria Savioli graduated in Sociology and specializes in Social Research Methods and Techniques. She has been working in the field of social statistics at Istat since 2000. Her research interests include well-being, educational poverty, social inequalities

Federico Segato is a postdoctoral researcher at the Center for Economic and Social Dynamics (University of Bergamo). He holds a PhD in Economics, with research on intergenerational mobility and inequality of opportunity. His expertise includes applied econometrics, machine learning, NLP and large language models for extracting data from complex unstructured texts.

Elisabetta Segre holds a PhD in Economics and has been a researcher at the Italian National Institute of Statistics (ISTAT) since 2010, where she specialises in policy evaluation. Her research interests include poverty analysis — with a focus on educational, economic, and energy poverty — inequality, and welfare.

Antonio Severoni, Technologist (Level II). Software Analyst and Developer. Holds a Master's degree in Educational Statistics and Research Methods from Sapienza University of Rome, and an Advanced Training Diploma in Cybersecurity and Cyber Rights Protection. Currently responsible for software development and database management in the INVALSI research area, Rome, Italy.

Georgia Solomonidou is a Scientific and Research Collaborator at European University Cyprus. She holds a PhD in Educational Research and an MSc in Educational Research Methodology from the University of Leicester. Her teaching and research focus on education sciences, and she has published in international journals and presented at conferences in Cyprus, Greece, and the UK.

Mara Soncin is a Senior Assistant Professor at the Department of Management, Economics and Industrial Engineering, Politecnico di Milano, Italy. She attained the PhD degree with a thesis on the evaluation of the performances of digital learning initiatives in Higher Education. Since 2015, she collaborates on research projects related to performance management in public administrations, with a specific focus on the educational sector. Her research interests also involve school management and leadership and the use of quantitative models for the evaluation of educational policies. She is also involved in teaching activities and training both in University and in post-lauream courses.

Francesca Storai is a researcher at INDIRE. Her research focuses on innovative training models for teachers and school leaders, school improvement planning, evaluation tools and educational innovation, with particular attention to processes that support organisational change in schools.

Rolf Strietholt is Associate Professor of Educational Sociology at Aarhus University, Denmark. His research focuses on international large-scale assessments, educational effectiveness, equity, comparative education, and the methodological development and secondary analysis of large-scale educational data.

Ioulia Televantou is Assistant Professor in Quantitative Research Methods in the Social Sciences at European University Cyprus. Her research spans large-scale assessment, motivational and emotional outcomes in the learning of mathematics and statistics, and at-risk populations. Her work appears in leading journals and has received international recognition through competitive awards and grants in educational research.

Mao (Mark) Tianzheng is a PhD student and Research Specialist. His academic and professional endeavors are rooted in the field of educational measurement and data analysis, with a particular focus on Item Response Theory (IRT) modeling. In his current role, Mark is engaged in a comprehensive study to evaluate the efficacy of various linking methods for IRT modeling. Specifically, he is exploring techniques that

facilitate the linking of a large number of independent calibrations. This includes an in-depth analysis of the simultaneous linking method proposed by Haberman and the Alignment method developed by Muthén et al. He is also investigating the potential of machine learning techniques to achieve similar linking objectives.

Kateryna Tkach, PhD, is an Assistant Professor in Applied Economics at the Gran Sasso Science Institute (L'Aquila, Italy). Her research focuses on regional disparities, socioeconomic inequalities, human capital and skills, forced migration, with particular attention to vulnerable groups, including women, immigrants, and refugees. ORCID ID: https://orcid.org/0000-0001-9407-6361.

Erika Daria Torello is a PhD student at the University of Jaén and has worked for more than twenty years in Italian primary school. Her research interests include gifted education, early identification of high potential, teacher training, and educational equity, with particular attention to teachers' perceptions and inclusive recognition practices.

David Torres Irribarra is an Associate Professor at the Pontificia Universidad Católica de Chile. His research areas include: the application of latent variable models to measurement contexts; the theoretical foundations of social sciences measurement (see his book "A Pragmatic Perspective of Measurement" (2021, Springer); and the use of digital technologies to enhance educational assessment.

Alessia Travaglini is a tenure-track researcher (RTT) at Università degli Studi Guglielmo Marconi in Rome. Her research focuses on pedagogy and teaching for inclusion**.**

Debora Tringali is a PhD student in Digital Transformation at the Pegaso Telematic University and an adjunct professor of psycho-pedagogical at the IUL Telematic University. She is the author of scientific contributions dedicated to general and social pedagogy, inclusion, emotional education, artistic languages and educational processes in school, museum and training contexts.

Teodora Erika Uberti is associate professor of Political economy at the Faculty of Political and Social Sciences of Università Cattolica del Sacro Cuore, where she teaches Economics and quantitative subjects. Her research interests focus on the study of social networks — their structure and the patterns of relationships among individuals — and how these patterns relate to socio-economic outcomes such as job satisfaction, innovation processes, and school success.

Valeria Maria Urbano is a Junior Assistant Professor at the Department of Management, Economics and Industrial Engineering at Politecnico di Milano. She earned her Ph.D. in Data Analytics and Decision Sciences from Politecnico di Milano in 2023. Her research focuses on public transportation management, healthcare systems, and the sustainable development of territories, employing data-driven methodologies to support decision-making and service innovation. She has experience in inter-organizational governance in the healthcare sector, gained through participation in European projects.

Cristian Usala is a PhD. Assistant Professor (Tenure Track) at the Department of Political and Social Sciences, University of Cagliari, and a researcher at the Centre for North-South Economic Research (CRENoS). His research focuses on social statistics, education, migration, microeconometrics, and public economics.

Giulia Vaiani is a second-year PhD student in Management Engineering at Politecnico di Milano. Her research covers educational inequality, ICT integration in schools, and socio-emotional skills. She is also involved in a Horizon Europe project on education quality and in a project with the Italian Ministry of Education and Merit on short-term teacher substitutions.

Benedetta Valpreda holds a master's degree in Statistics and is an Evaluation Specialist at the Evaluation Lab of the Fondo per la Repubblica Digitale, which supports digital training projects aimed at fostering the country's technological transition. Her research focuses on evaluating the impact of the initiatives supported, with the aim of identifying the most effective approaches to reducing the digital divide.

Loris Vergolini is an Associate Professor at the University of Bologna from 2024 and was previously a researcher, also at Bologna, from 2021 to 2024 and at FBK-IRVAPP from 2009 to 2021. He conducts

research in the field of policy evaluation, with a focus on policies to widen participation in higher education. His research interests include the study of educational inequalities, social cohesion and the cultural aspects of ageing. Since 2024, Loris Vergolini has been working with the Impact Assessment Office of the Senate of the Republic.

Karine Verschueren is full professor at School Psychology and Development in Context, KU Leuven. Her research focuses on how students' academic, motivational, and psychosocial development is shaped by social relationships and interactions at school. She studies these developmental processes among general student populations and high-ability students, with special attention for educational transitions. For her publication record, see: https://orcid.org/0000-0003-2172-1424.

Iván Vicente Carrión is an Assistant Professor in the Department of Economic Analysis at the University of Valencia. His research addresses the economics of education and educational inequality, peer effects, and the impact of AI on productivity.

Daniele Vidoni holds a PhD in Economics of Education from Boston University and is a policy official at the Directorate-General for Education, Youth, Sport and Culture (DG EAC), European Commission. His work focuses on education policy evaluation, large-scale assessment, and the use of evidence in education policy-making across European school systems.

Giada Viotti is a PhD candidate in Educational and Social Research at Roma Tre University, supported by a doctoral scholarship co-funded by INVALSI. She collaborates with INVALSI on ILSAs, with particular attention to TIMSS data, learning outcomes, teaching practices, and equity in education. She holds degrees in Social Psychology, Pedagogy and Education Sciences, and Primary Education.

Matthias von Davier is Monan Professor in Education. Executive Director, TIMSS & PIRLS International Study Center Dr. von Davier's research focuses on developing psychometric models for analyzing data from complex item and respondent samples and on integrating diagnostic procedures into these methods. His areas of expertise includes topics such as item response theory, latent class analysis, classification and mixture distribution models, diagnostic models, computational statistics, person-fit, item-fit, and model checking, as well as hierarchical extension of models for categorical data analysis, and the analytical methodologies used in large scale educational surveys.

Jan–Philipp Wagner is a Research Associate at the Department of Education at the University of Bath. After previous work at the International Association for the Evaluation of Educational Achievement (IEA) in Amsterdam, he currently works as Study Coordinator for the International Civic and Citizenship Education Study (ICCS) 2027.

Giulietta Zanga is a postdoctoral researcher at the University of Milano. Her research interests include educational inequalities and gender inequalities. She recently defended her PhD thesis, which investigated teachers' social networks and how they influence their stereotypes concerning students with different ascribed characteristics.

Benjamín Zavala is a student of Sociology at the Department of Sociology, Universidad de Chile. His research interest is related to citizenship education and youth participation.

Via Ippolito Nievo, 35 – 00153 – Tel. 06941851
Mail: protocollo@INVALSI.it PEC: protocollo.INVALSI@legamail.it Website: www. INVALSI.it
C.F. 92000450582