INVALSI

datiinvalsi
I dati del e per il sistema educativo: strumenti per la ricerca e la didattica

X SEMINAR

"Data from and for educational systems: tool for research and teaching"

Thomas George Webster, A Dame's School, 1845

Rome, 19 - 21 November 2025

Under the patronage of

espanet ITALIA

Istat
Istituto Nazionale di Statistica

SIRD
Società Italiana di Ricerca Didattica

INDEX

INTRODUCTION

The Seminar "Data from and for educational systems: tools for research and teaching", now at its eighth edition, has become in recent years an opportunity for meeting and discussion on the use of National Surveys INVALSI results and, in general, on the world of evaluation and school. Evaluation does not only mean standardized monitoring of learning levels, but also evaluation and comparison of its functions and potential in relation to the practices and tools through which the school system is able to carry out its educational and formative mission. This year, keynotes of national and international experts were organized on issues related to the evaluation of the education system and the use of data to support school policies.

One of the objectives of the seminar was to bring the world of scientific research and schools closer together in order to build a privileged and concrete space in which enriching the debate by sharing ideas and experiences between education and research stakeholders. The variety of topics improve a multidisciplinary approach to evaluation in the educational and school field, trying to give an account of the contribution that schools can provide to society by making possible the development of knowledge and skills.

The Seminar 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, Alessandro Gaeta, Paola Giangiacomo, Patrizia Giannantoni, 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, Massimo Smiraglio, Daniele Rowlett, Antonio Severoni, Agnese Spoladore, Valeria F. Tortora.

Contacts: uff.statistico@INVALSI.it

ABSTRACT

SESSION 5. ASSESSING DIGITAL SKILLS AND ASSESSING LEARNING WITH DIGITAL: NEW CHALLENGES FOR SCHOOLS

ORGANIZER: INVALSI - UNIVERSITY OF BERGAMO
COORDINATOR: PAOLO BARABANTI – MARCO GIGANTI
NOVEMBER 19TH: 2.00 P.M. – 4.15 P.M. {AULA MAGNA – RESEARCH 1}

From Frameworks to Classrooms: Digital Competences in Action. Insights from the INVALSI 2023-24 Teacher Questionnaire

Marco Giganti – Paolo Barabanti

Introduction

The integration of digital technologies in schools is now a fundamental part of teaching and learning, rather than an optional extra. Numerous international policy documents and guidelines, particularly those from the European Union, emphasise the urgent need to promote widespread digital literacy among students and teachers alike (European Commission, 2020; UNESCO, 2021). Initially defined mainly in technical terms, the concept of digital competence has evolved into a broader construct that encompasses the critical, ethical and pedagogical dimensions of living and learning in the digital age (Ferrari, 2013; Redecker, 2017).

Within the European Union, the DigComp2.2 framework (for citizens), the DigCompEdu framework (for educators) and the DigCompOrg framework (for educational organisations) are key tools for organising the digital competences necessary for active citizenship and innovative education. In Italy, the National Digital School Plan (PNSD), the goals of the 2030 Agenda and the measures under the National Recovery and Resilience Plan (PNRR), particularly Mission 4 – Education and Research, have adopted the principles set out in these documents. They promote training initiatives for teachers, the development of digital infrastructure and the use of technology to innovate teaching methods.

However, a significant gap remains between the theoretical framework and actual practices. There is a lack of empirical studies that systematically assess Italian teachers' knowledge of, and ability to apply, digital competence frameworks, as well as their perceptions of students' digital competence. Similarly, the extent to which technologies are used in education and the ability to integrate them critically and reflectively continues to be a key research area, particularly given the diversity of teacher profiles and educational contexts across schools and regions.

Objectives and Purpose

This study aims to partially fill this knowledge gap by offering an exploratory analysis based on data from the INVALSI 2023/24 Teacher Questionnaire, administered during the national assessment cycle to teachers whose classes were selected as part of the national sample. The general aim is to investigate Italian teachers' enacted digital competences and their perceptions of students' competences, focusing on three dimensions: 1) teachers' knowledge and use of key European frameworks (DigCompEdu, DigComp2.2, DigCompOrg); 2) perceptions of students' digital competences; 3) teaching practices involving the use of digital technologies.

Based on these objectives, the following research hypotheses were formulated:

- H1: Teachers with more years of service and those working in upper secondary schools are more familiar with European digital competence frameworks.
- H2: Teachers' perceptions of students' digital competences vary significantly by school level and subject taught.
- H3: Educational use of technologies is positively correlated with teachers' self-reported familiarity with European frameworks, and this depends on their age and years of experience.
- H4: Integration of digital technologies in teaching practices is higher among STEM teachers and in lower-level schools.

Data Used

The data come from the INVALSI Teacher Questionnaire, which is one of the Integrative Surveys administered to teachers with a sample class in the National Assessment. A total of 8,001 teachers were surveyed. The sample includes teachers from primary school (grades 2 and 5), lower secondary school (grade 8) and upper secondary school (grades 10 and 13), and is representative of the tested subjects (Italian, mathematics and English). Stratifying the sample allows analysis of variables in relation to macro-geographic areas (North-West, North-East, Centre, South and Islands) and years of teaching experience.
The analysed sections of the questionnaire refer to:

- knowledge and use of European digital competence frameworks (DigComp2.2, DigCompEdu, DigCompOrg).
- Teachers' perception of students' digital competences.
- Digital teaching practices (e.g., use of IWBs, e-learning platforms, educational software and apps).
- How digital competences are addressed at school.

Method

The study takes a quantitative approach, which is divided into several phases. First, descriptive analyses were conducted to provide an overview of the main variables' absolute and relative frequencies. Secondly, contingency tables and chi-square tests were employed to examine the relationship between framework knowledge and the independent variables (school level, subject and years of service). Next, an exploratory factor analysis was conducted on a set of items related to the use of digital technologies in teaching. The aim was to construct a synthetic index of 'digital integration in teaching', which was then used as a dependent variable in subsequent analyses. This index was then compared across categorical variables (school level and subject) to highlight potential group differences. The results presented here focus on the most relevant emerging patterns, with further multivariate analyses planned for future studies.

Results

Despite the consolidation of theoretical and policy references, the results suggest that there are still significant gaps in Italian teachers' knowledge of, and use of, European frameworks. Similarly, the integration of digital technologies in teaching remains uneven, with substantial differences relating to school level, subject area, and years of service. Key findings include:

- Only a minority of teachers reported reading or working with the European digital competence documents. Nevertheless, there appears to be consistency between teachers' professional profiles and the frameworks used. DigCompEdu is the most cited among educators, while DigComp2.2 is known by a smaller group.
- Contingency analyses show a significant association between school level and knowledge of the frameworks: teachers at upper secondary level and those with fewer years of service report higher familiarity.
- Teachers' perceptions of students' digital competences are mixed: while many consider them 'adequate for the grade level', a notable proportion rate them as 'very low' or 'inadequate'.
- The way digital competences are addressed at school varies too: most teachers describe them as cross-curricular, but only a minority report that their schools treat them as a standalone subject.
- The synthetic index on digital integration in teaching shows higher average values among English teachers, in lower-level schools and among less experienced teachers, confirming a certain polarisation in the adoption of digital practices.

Keywords: Digital competences, EU DigComp frameworks, INVALSI Teacher Questionnaire, teaching practices

Profiling ICT Use and Perception in European Schools: A Two-Level Latent Class Analysis

Giulia Vaiani – Mara Soncin – Tommaso Agasisti

The integration of Information and Communication Technologies (ICT) in education is increasingly recognized as a key driver for enhancing teaching practices, supporting student learning, and fostering digital skills essential for the 21st century. While access to digital tools has expanded significantly in recent years, the effective use of ICT in schools depends not only on infrastructural availability but also on the attitudes, beliefs, and digital readiness of the various educational stakeholders. A positive orientation

towards ICT is often associated with greater willingness to innovate pedagogical practices and to engage with technology in meaningful ways. Understanding how different actors experience and perceive ICT in school settings is therefore essential for designing policies and interventions that support effective digital integration.
In this article, we investigate the use and availability of ICT in education through a two-level latent class analysis based on survey data collected from primary and secondary schools across four European countries (Italy, Belgium, Hungary, and the Netherlands). We begin by analyzing how students and teachers can be grouped on the basis of ICT use and accessibility, considering their nested nature within schools. In the first step we examine the relationships between student and teacher classes within the same school context. We then explore cross-country heterogeneity in the distribution and interpretation of these latent classes. In the second step, we explore teachers' attitudes and perceptions towards ICT use in education, and investigate how these relate to the previously identified usage classes. Finally, we analyze whether differences in student perceptions can be interpreted as outcomes associated with class membership.
The results of this analysis offer valuable insights into the complex dynamics of ICT integration in schools, highlighting patterns of use and perception across different educational roles and national contexts. By uncovering latent classes and their interrelations, the study contributes to a more nuanced understanding of how digital technologies are experienced within the school ecosystem. These findings carry important practical implications for researchers, policymakers, and educational practitioners aiming to promote more effective and inclusive strategies for digital transformation in education.

Keywords: ICT, Teachers Perception, Latent class analysis, Cross-country comparison

A Community of Practice Focused on the Development of Adaptive Assessments for Estimating Mathematical Proficiency at the Fifth-grade Level

Emanuela Botta – Stefania Pozio

Introduction

In the field of research on objective assessment methods of academic achievement, adaptive test models are now widespread compared to traditional linear formats. The advantages of such models lie primarily in their ability to provide highly accurate estimates of each student's ability (Weiss, 1985; Weiss & Kingsbury, 1984; Hambleton, Swaminathan & Rogers, 1991; Botta, 2019, 2021a, 2021b; Yamamoto, Shin & Khorramdel, 2018). In classical linear tests, which are centered around the population mean, the informational value and measurement precision significantly decrease as one moves away from the center of the score distribution. As a result, the estimation of student abilities using traditional linear tests is often based on a very limited set of items, which rarely supports meaningful feedback on students' strengths and weaknesses. For instance, in the 2018 and 2022 OECD PISA assessments—first in reading, then in mathematics—an adaptive testing model was adopted specifically to "improve the precision of measuring student performance at the upper and lower ends of the proficiency distribution (i.e., among high- and low-performing students)" (OECD, 2023). In adaptive models, rather than assigning all students the same fixed set of items, the items administered are selected dynamically throughout the test based on each student's performance at various stages.

Purpose, Objectives, and Research Hypotheses

The AMAT project, *Development of a Battery of Valid Questions for Assessing Mathematical Competencies in the Classroom with Adaptive Test*, funded and conducted as part of the broader UNIT–T European Civic Teacher Academy initiative, aimed to establish a community of practice focused on equipping teachers with specific skills in mathematics assessment—from the construction of valid, fair, and effective tests to the use of formative and educational assessment practices. The project took place from October 2023 to May 2025.
Research Questions:

1. Is it possible to build an open and freely accessible item bank for the development of adaptive tests to be used by schools?
2. How can schools design and implement adaptive mathematics assessments?
3. Will the automatic feedback model be effective and well-received by students and teachers?

Specific Objectives:

1. To develop an MST (Multistage Testing) adaptive test for assessing mathematical competencies.

2. To construct mathematics items aligned with the assessment framework and specific learning objectives.
3. To estimate item difficulty using various models.
4. To identify each student's strengths and weaknesses in relation to the assessment framework.

DataUsed

The framework used for designing the test was based on the INVALSI Mathematics Framework (INVALSI,2018) and the TIMSS 2023 Framework.
Additionally, to establish a reference base for calibrating item difficulty, several items from the INVALSI grade 5 assessments were selected and included in the test.
The AMAT test was pretested at grade 6 level on a sample of 674 students. Following the pretest, the test was calibrated using the one-parameter IRT model. Item functioning and overall test performance were analyzed, and necessary modifications were made to produce the final version, which was administered between April and May 2025 to a sample of over 400 fifth-grade students.

Method

The research design was based on the creation of a community of practice, following the research-training approach. This community included French teachers from the Marseille area-coordinated by colleagues from Aix-Marseille University-and Italian teachers from various comprehensive schools in Rome.
Within this community, an MST (Multistage Testing) model was developed, now available for use by teachers in their schools at the end of grade 5 or the beginning of grade 6.
Teachers collaborated in developing an item bank suitable for constructing this type of test. Each participating school had the opportunity to engage in the different stages of test development—from defining the assessment framework, to constructing individual items, and finally assembling the complete test. In some classes, feedback based on formative assessment principles was also piloted (Hattie, 2016; Hattie & Timperley, 2007; Black & Wiliam, 2009; Botta, 2022, 2023; Lilley & Barker, 2007; Lilley, Barker & Britton, 2005).
The teachers in the community of practice, divided into working groups, developed the test items starting from a reference framework defined by Italian and French researchers. The items were reviewed and revised by the research team, which also handled the translation.
To estimate item difficulty in the initial version of the test, an adaptation of the Angoff method was used (Angoff, 1984; Cizek & Bunch, 2007; Lane et al., 2016).
First, the teachers who authored the items provided a difficulty estimate on a scale from 1 to 5, based on their teaching and assessment experience. In the second phase, researchers independently provided their own difficulty estimates using the same scale. Finally, the research team reviewed all estimates and assigned each item a final difficulty level, considering the previously expressed evaluations.
A three-level adaptive test was developed, consisting of a base module (level 1) administered to all students, and two adaptive modules (levels 2 and 3), assigned based on students' estimated ability levels—low, medium, or high.
Each student responds to items from three modules, one at each stage of the test. The figure, for example, highlights the path followed by students with the lowest ability level, who receive easier modules.

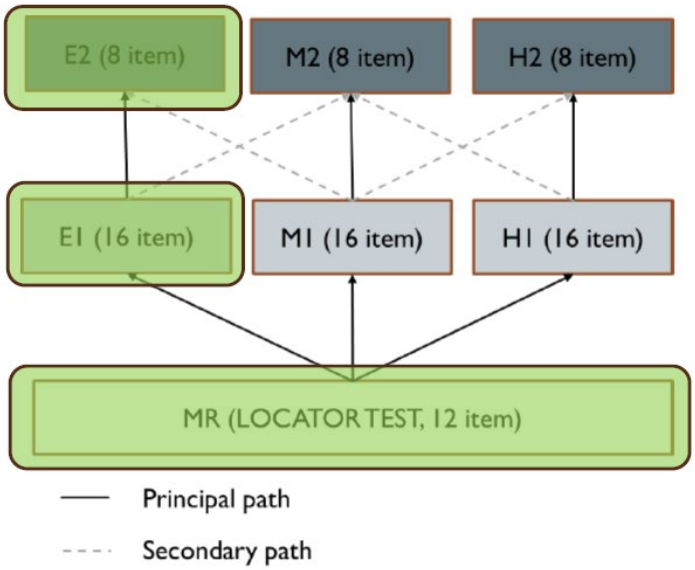


Figure 1 - Structure of the AMAT Test

Results

As part of the project, a total of 201 items were developed, 84 of which were selected and used in the first version of the test and subsequently pretested. Of these, only 5 showed psychometric issues significant enough to warrant removal. A few other items were modified to improve their performance.

It is worth noting that in the AMAT test, the discrimination index of items in the intermediate and final modules — particularly in the easier modules administered to a small number of students —was quite low, sometimes even negative. One possible explanation for this phenomenon, even for items that otherwise functioned well, is that the test was administered directly in adaptive format: modules at levels 2 and 3 were only given to students with very similar ability levels. The minimal variation in respondents' abilities made it difficult to accurately estimate item discrimination. For example, the easy level-2 module, E2, was taken only by students with low proficiency. This issue was not observed in the base module, MR, since it was administered to all students, who showed a broad range of ability levels overall.

Calibration and scaling analyses were conducted for the entire test. The information provided by the test at different levels of theta was measured using the Fisher Information Function (Efron & Johnstone, 1990), which estimates the amount of information the test yields at each ability level. The highest observed information value, based on real data, was approximately 17 around theta = 0, compared to a theoretical maximum of 21.

The application of the Angoff method to estimate item difficulty—although aligned with the project's goals—likely represented its weakest point. In 61% of the cases, the difficulty classification assigned by teachers and experts was incorrect, especially in the geometry domain. In the "Numbers" domain, 30 out of 56 items were misclassified, with over- and underestimations evenly distributed. In contrast, in the "Geometry" domain, 21 out of 28 items were misclassified, and in 70% of those cases, items were judged to be easier than they actually turned out to be.

The test proved effective in estimating student abilities and demonstrated good adaptive behavior. The teachers involved in the community of practice expressed great enthusiasm for the project, which is expected to continue beyond its official end date.

Consistent with the literature, adaptive assessments appear more accurate and more suitable than linear tests for formative assessment as well as for system-level evaluation. They effectively address the need for a model that accommodates all students, including those with special educational needs.

Keywords: adaptive tests, formative assessment, mathematics, MST testing, feedback

Grade 5 mathematics test in digital and paper format: What do students think?

Laura Montagnoli – Francesca Ferrara

Introduction

The gradual transition from standardized paper-based to computer-based assessment, as is already the case in the OECD-PISA and IEA-TIMSS surveys and in many national surveys, is driven by a number of advantages, including: the comparability of results over time; centralized, thus objective and faster, correction, which offers teachers the opportunity for targeted interventions (McClelland & Cuevas, 2020); the inclusion of special features in the test, such as dynamic geometry tools (Threlfall et al., 2007); the possibility of detecting timing and proposing adaptive tests (Noyes & Garland, 2008); and savings in terms of money. However, the transition to digital assessment also has some disadvantages, e.g.: fatigue due to reading from the screen (Jeong, 2012); fragmented reading of the test, when it is not possible to see substantial parts of a question in one glance; the need for paper support for the performance of some items, with the consequent increase in the load in terms of working memory (Russell et al., 2003).

The transition is the subject of study in some research works that not only examine the validity and comparability of tests by analyzing the outcomes but also focus on the cognitive processes elicited by items offered in a different modality, i.e. paper and pencil or digitally (Threlfall et al., 2007). In these studies, items are typically exploited that were created for a paper-based test and were either migrated (with minimal changes) or transformed (with more impactful changes due to a different mode of interaction) to be included in a digital test (Grapin & Sayac, 2022), or items that are presented in the two different modes are compared (Lemmo, 2021).

In this paper, we focus our attention on the transition to digital tests at grade 5 in Italy, where up to now the assessment mode has always been paper-based. For this purpose, we consider some questions constructed for the digital format and afterwards also administered in paper format and the processes by computer-based resolution as opposed to paper-based, as well as advantages and disadvantages that seem to emerge.

Subject And Research Aims And Hypotheses

The work intends to reflect on the way in which grade 5 students handle a computer-based mathematics test and how this differs from a paper test.

At the same time, it intends to investigate how the presence of a sheet of paper can support the students in solving the items, given their habit of working with paper, and whether their preference falls on one or the other method of administering the test. To this end, we consider both what students state while reflecting aloud on the preference of one or the other mode and through the analysis of their written protocols, which consist of the productions on the sheets of paper they had at their disposal when taking the test in digital mode.

The main aim is to investigate the following questions:

- while students are taking the test digitally, do they use the paper they have been given? If yes, for which tasks do they use it?
- What do they think about the advantages and disadvantages of the digital test compared to the paper test?

The answers we can give to these questions offer insights for researchers, teachers and designers of INVALSI questions.

Underlying the study is the idea that the paper provided can play an important role especially in solving those items that require calculations or the manipulation of figures, which students might therefore prefer to tackle in paper mode.

Data And Methodology

In April and May 2025, INVALSI involved a sample of 240 grade 5 classes in the administration of a mathematics test in digital mode, using tablets or PCs. Students were also provided with a sheet of paper as a possible support for calculations, reasoning and representations. The items used for the test were produced by a group of experts in order to be included in a future national test carried out digitally and were compiled in twenty forms, with partial overlap. The items cover all the content domains of the Mathematics Reference Framework used by INVALSI (*Numbers*, *Space and figures*, *Relations and functions*, *Data and predictions*) and all the cognitive dimensions into which the Competence Development Goals are grouped (*Knowing*, *Problem solving*, *Arguing*). They are also representative of the main Learning Goals of the 2012 National Curriculum guidelines.

From the 240 classes involved, 56 students from a total of 7 classes were randomly drawn and, after the administration of the test, took part in an interview in pairs. During the interview, each pair was presented with 4 items printed on paper that had already been administered in the test in digital format. These items were selected from a set of 8 items, all multiple-choice, covering the four content domains and the three cognitive dimensions.
The interview focused on the following aspects:

- the preference between the paper and digital modes in relation to what the items required and whether or not additional elements were needed to support the reasoning;
- the general considerations regarding carrying out the items in the two different modes.

In addition, the additional sheet provided during the digitally conducted test was collected for 5 of the 7 classes involved in the interview.

Results

A first general result concerns the use of the sheet of paper provided during the digital test, which is exploited by just over 86% of students. Beyond this information, it is interesting to focus on what is found in writing and to compare it with the statements of the interviewees. In particular, considering all the sheets collected from the classes and specifically those of the interviewees, we see that the main uses of the sheet of paper are related to:

- calculation (multiplications in column between natural factors and decimals, multiplications in line between natural numbers, divisions between natural numbers, additions in column between natural addends and decimals, subtractions mainly between natural numbers);
- the reproduction of figures both to deal with geometric tasks, such as the search for properties of polygons, and to search for regularities in sequences;
- the listing of sequences of numbers or of times to explore properties of natural numbers (such as being a multiple) or regularities, to support e.g. direct proportionality reasoning;
- the annotation of information or values found from the question or numerical answers useful for reasoning.

From a qualitative point of view, the interviews confirm these types of use of the sheet of paper and highlight justifications, such as: "*because you have to do the calculation and if you can't do it in your head*", "*also to do drawings, when needed*". On the other hand, one student mentions the need to use the paper for procedures related to the use of formulas: "*in a geometry exercises you have to do the perimeter, then the area, then the inverse formulas, you have to use the paper to write it down, so that if you write it down you remember it better*", shifting the focus to the role of writing to support memory. Further considerations emerge concerning the need or possibility of:

- gaining more awareness of one's own thinking: "*if you don't know an exercise, you can do operations, instead with the computer… we won't know what I had to do in my mind*";
- organising information and solving procedures: "*because I also find myself making diagrams to understand*";
- activating control processes: "*I would have written to control*".

Not everyone, however, uses the paper. Some state that they do not need it either for calculation ("*if you have done maths and know the basic calculations, you can do them in your head*") or for drawing figures ("*I just imagine them*", "*I tried to see them with my finger*", "*drawing them in my mind*") or even for counting processes.
46 participants declare that an imaginative process is sufficient by providing supporting reasons, while 34 entries emphasise the need or importance of interacting directly with the question ("*I would like to draw them, put them in the figure, try to do them*", "*for me here it is better on the paper because you can also touch, you can count*"), someone even for the *Space and figures* domain ("*yes, I also prefer arithmetic on the tablet because there you do the calculation on a sheet, instead geometry you could also do on top of the figures or next to them*").
It is therefore interesting to examine the reasons why respondents prefer to take the test in paper or digital mode. First of all, there are more arguments in favour of taking the test digitally than on paper (in the interviews, 156 preferences were expressed for the use of a device, 87 arguments that there is no difference between the two modes, 78 preferences for the use of paper).
Besides general issues, which mainly concern the habit of using paper and the different time management in one and the other case, certain aspects (from the most to the least recurring) frequently arise:

- the advantages offered by the screen and the functions of the device ("*you can zoom in and go and see better*", "*you can check which questions you have answered, you can mark the ones you need to look at again*");
- the insecurity in using the device, perhaps due to a lack of familiarity ("*you risk opening something, moving something*");
- the ease of changing answers digitally, thus being able to take advantage of an exploratory approach and avoiding erasure marks ("*I marked it, you could correct it later anyway*", "*it's also easier to mark if you got the answer wrong because you erase it and put it back, instead then at some point if you realise you got it wrong you make a mess of it*");
- the better readability and less fatigue offered by paper ("*I look better on paper*", "*paper is less tiring*");
- the positive emotions associated with the use of the devices ("*it's nicer on the computer, a child is excited on the computer*");
- the difficulty in repeatedly switching from the screen to paper, in the case of using the backing paper ("*taking my eyes off the image to write on the paper*").

In conclusion, not only do useful insights emerge from this analysis to observe the processes that can be stimulated by solving mathematics items in a digital format, but also a descriptive picture of key elements when facing a digital test emerges, such as the following: the importance of remembering, insecurities or emotions in front of a screen, the need to activate control processes, and awareness of the tools useful when dealing with certain mathematics questions.

Keywords: CBT, paper and pencil, mathematics assessment, student preference

Computers and Tablets: An Exploratory Study on the Impact of the Possibilities Offered by These Digital Tools

Cintia Scafa Urbaez Vilchez

Introduction

For the past twenty years, the use of digital devices for the administration of standardized tests has been a topic of great interest for international assessment institutions (OECD, IEA) and academic communities. In the assessment literature, the term *Computer-Based Testing* (CBT) refers to the use of digital tools (mainly personal computers and tablets) for test delivery. The widespread and intensive use of digital tools in various socioeconomic, cultural, and everyday contexts has undoubtedly guided many researchers toward introducing these systems in educational settings. Moreover, using digital media in this domain offers advantages in terms of cost, logistics, and learning. Compared to its paper-based counterpart (Paper-Based Testing, PBT, or Paper and Pencil Testing, PPT), the CBT modality allows for a reduction in costs related to printing, transportation, and large-scale distribution of tests to schools [Fishbein et al., 2018; Lynch, 2022; Poggio et al., 2005]. Furthermore, CBT enables the collection of more data on tests (such as time taken to complete the test, attempts made by students, etc.), and results can be returned more quickly and securely thanks to dedicated software that ensures data integrity and security. Additionally, multimedia features of digital tools make it possible to introduce response interactions that are different from those used in paper-based tests [Redecker & Johannessen, 2013; Sibberns, 2020].

As already mentioned in the introduction, the integration of technology into assessment also presents challenges, such as determining whether digital and paper tests are comparable—that is, whether they meet validity and fairness standards. *Validity* refers to how accurately a test measures the intended construct, while *fairness* ensures the test provides equal conditions for all students. Ripley (2009) identifies two approaches for transitioning from PBT to CBT: *migratory* and *transformative*. In the migratory strategy, "paper-based tests are migrated to a screen-based environment, but their quality remains unchanged" [Ripley, 2009, p. 87, author's translation]. In the transformative strategy, however, "the test is constructed in a way that brings (or at least reflects minimal) innovation in its development and learning potential" [Ripley, 2009, p. 87, author's translation].

As previously noted, digital tools allow for the inclusion of new item types that leverage multimedia features. In the transformative strategy, tests consist of technologically innovative items that assess skills not measurable through paper-based formats, such as the ability to interact with information systems in a

virtual and three-dimensional context. In this regard, Bennett (2015) describes three generations of technology-based assessment:

1. One very similar to Ripley's (2009) migratory strategy.
2. A second, in which questions are transformed to make an "initial attempt to measure new constructs, starting to change what is being assessed" [Bennett, 2015, p. 371, author's translation].
3. A third that aligns with Ripley's (2009) transformative approach.

Another key academic topic concerns the comparison of student behavior, actions, and problem-solving processes across the two delivery modes. Some researchers [Russell et al., 2003; Threlfall et al., 2007; Lemmo, 2023] claim that digital tools possess *affordances*, defined as the effect of the interaction between users and the computer interface on their responses [Threlfall et al., 2007]. In other words, affordances are the possibilities for action offered by a digital interface in relation to the user's abilities. Threlfall and colleagues (2007) highlighted differences in outcomes between the two formats. For certain interactions, such as ordering digits, numbers, or measurements to form a value or sequence with specific characteristics, they argue that digital affordances (in this case, of the computer) are greater than those of the paper-based version, as they allow students to make more attempts and reduce cognitive load [Sweller, 1994]. In contrast, Lemmo (2023) shows that students' problem-solving processes are similar in both formats, although the resources used to answer questions may vary.

Research Aim, Objectives, and Hypotheses

Building on the work of Threlfall and colleagues (2007), this study focuses on the impact of presenting digital items through two types of digital devices, computers and tablets, on students' problem-solving activity. We also ask whether and how the *affordances* of two different interaction types have a varying influence depending on the device used.

This paper focuses on how fifth-grade students interact with digital math items using two interaction formats: *drag and drop* (element association via dragging) and *inline choice* (dropdown menu), on both computers and tablets. These two response interactions do not alter the mathematical content of the items but affect the user experience.

Examples of drag-and-drop items include placing numbers on a number line or filling in missing values in a table. Dropdown menus may be used, for instance, to complete a text with missing data or to fill in a table. Based on the literature, we situate our test in the second generation described by Bennett (2015), between the migratory and transformative approaches, since we selected response interactions typical of digital environments.

By *impact*, we mean whether students experience interruptions during problem solving, take longer to respond, and/or require technical assistance to continue.

Data Used

To carry out our study, we created an observation grid to count, for each interaction type and each difficulty encountered, how many students requested assistance or became stuck during the task.

For drag and drop, obstacles included: difficulty selecting the element, holding and dragging it to the target space, or releasing it in the correct location. We recorded the number of help requests for each obstacle and the related item.

Methodology

Our study took place during the trial of the fifth-grade digital math test conducted by INVALSI from March 31 to June 6, 2025, across various primary schools throughout Italy. The goal is to introduce CBT for this grade level.

The test, accessible via the TAO platform, included questions across four content domains (Numbers, Space and Shapes, Data and Forecasts, and Relations and Functions), three dimensions (Knowing, Problem Solving, and Arguing), and addressed curriculum-aligned topics. Each item included instructions, a video tutorial, and a simulation made available before test administration.

While dropdown menus are used by clicking on a small arrow, TAO allows drag-and-drop interaction either by dragging the item or by clicking both on the item and the target space.

The author conducted observations in six schools: three in Perugia (Center), two in Naples (South), and one in Modena (North), totaling eleven classes. Two schools used tablets; of the remaining four, three had laptops with mouse, and one had laptops without mouse.

Thus, we analyzed three conditions: tablets, laptops with mouse, and laptops without mouse. During the test, the author recorded students' interactions with the platform, technical help requests, and difficulties that hindered or slowed down test completion.

Results and Discussion

Results show that in the laptop-with-mouse condition, there were no significant difficulties. However, 12% of tablet users had trouble with drag and drop, specifically dragging the selected element and changing their answer. Students reported that "the only way to respond was to click the element and then the blank space," while with a mouse, dragging and dropping seemed easier.
Additionally, about 21% of students using laptops without mouse struggled with this interaction.
For dropdown menus, data indicate no significant issues across the three conditions. However, 0.07% of tablet users encountered problems due to the virtual keyboard covering half the screen, which distracted from problem solving and forced students to open and close the menu repeatedly to reread the question.
In conclusion, we observed slight differences among the three conditions: laptop with mouse, laptop without mouse, and tablet. For dropdown menus, larger screens may reduce discomfort, as the menu covers less of the question.
For drag and drop, no major differences emerged between the two laptop conditions, except that 21% of students using laptops without a mouse had more difficulty dragging items, possibly due to unfamiliarity with touchpad use, which requires coordinating both hands.
Tablet use also posed challenges: pressing and dragging elements with a finger across the screen appeared difficult for students. Nonetheless, in all three conditions, we observed no cases of students giving up on the test due to frustration.
A potential future development of this research could be to compare these qualitative data with platform performance data, to assess whether students' difficulties with the two interaction types influenced not only the response times for individual items but also for the test as a whole.

Keywords: CBT, impact, drag and drop, dropdown menus

The INVALSI DigComp digital competence assessment: a data analysis to understand student proficiency profiles and support didactics

Paola Giangiacomo - Monica Papini - Valeria Fortunata Tortora

Theoretical Framework: The pervasive integration of digital technology into schools necessitates a critical reflection not only on technical mastery but also on the ability to teach and learn in an increasingly technological environment. Digital competence is now an essential part of a continuum that includes traditional literacy, requiring students to interpret, produce, and evaluate digital content for active citizenship. The school system faces the dual challenge of promoting critical and conscious use of technology among students and supporting teachers in developing innovative teaching and assessment practices. In this context, the urgent need to strengthen digital competencies and bridge the gap between students and teachers highlights the centrality of valid tools for systemic and standardized assessment of digital competencies, such as INVALSI surveys.
Research Objectives/Questions: This contribution aims to address new challenges for schools by using INVALSI data to shed light on the development of digital competencies. Our objectives are: 1) To analyze the levels and profiles of digital competence among Italian students through INVALSI data, in reference to European frameworks (DigComp), to understand their proficiency and any critical areas. 2) To explore the relationships between digital competence profiles and individual variables (e.g., gender, results in other subjects) and contextual variables (e.g., geographical area, school type), to identify inequalities and gaps. Research questions include: "How are Italian students distributed across different levels of digital competence according to INVALSI tests?", "Are there specific and distinct digital competence profiles among students?", and "What implications can be drawn from these data for teaching innovation and teacher training in the context of bridging digital divides?".
How INVALSI Data Will Be Used: For this research, data from recent national INVALSI surveys on digital competencies (DIGCOMP) will be utilized. These large-scale collected data allow for a granular analysis of student proficiency levels across the various areas of digital competence outlined by the DigComp framework. Descriptive and inferential statistical analysis techniques will be employed to delineate competence profiles. Subsequently, regression analyses will be conducted to examine correlations between these profiles and individual and contextual variables available in the INVALSI datasets, in order to identify factors influencing digital competence acquisition and existing disparities.

Results: The expected results will outline a detailed picture of the levels and profiles of digital competence among Italian students, highlighting areas of strength and critical areas. We anticipate identifying specific student profiles, some of which may show significant gaps in certain digital competence areas, potentially correlated with socioeconomic or territorial variables. These profiles can shed light on the digital divide not only among students but also in relation to the resources and teaching practices present in school contexts. The results will contribute to a better understanding of the training needs for both students and teachers, providing an empirical basis for developing targeted training programs and for the effective integration of digital technologies in teaching and learning assessment.
Relevance for the Seminar: This contribution is highly relevant for the Seminar "X Seminario “I dati del e per il sistema educativo: strumenti per la ricerca e la didattica”," as it directly responds to the call to reflect on the definition and assessment of digital competencies and to present research based on national surveys regarding their levels and profiles. The analysis of INVALSI data offers an evidence-based perspective on the challenges schools face in the era of technological transition. The proposed research provides useful tools to monitor progress, identify gaps, and support the planning of educational interventions and teacher training, consistent with the need to integrate digital technologies into teaching and assessment, thereby promoting quality and inclusive education.

Keywords: Digital competencies, Assessment, INVALSI data, Digital didactics

Digital Learning Spillovers in Education. Why Additional Digital Training Can Backfire

Massimiliano Bratti - Federica Dalponte

Introduction
Digital technologies have become increasingly pervasive across everyday life, professional environments, and educational settings, fundamentally transforming the way individuals interact, work, and learn (UNESCO, 2023; World Economic Forum, 2025). It is therefore increasingly crucial for students to be equipped with robust digital skills that allow them to engage responsibly with digital devices in everyday life, leverage them effectively for learning, and position themselves competitively for emerging job opportunities. This is even more important given the intertwined relationship between socioeconomic inequalities and digital skill disparities, which are found to mutually reinforce each other (Van Dijk, 2020). Theoretically, providing digital skills to students could have both positive and negative effects on their academic performance. On one side, improved digital literacy might enhance new knowledge acquisition, critical online content evaluation, and more purposeful technology use for learning. Conversely, increased familiarity with digital devices could lead to higher non-educational device usage, potentially distracting students and reducing study time. There is an extensive body of literature investigating the effects of digital devices use or availability on academic performance, finding mixed results (e.g., Carter et al., 2017; Cristia et al., 2017; Fairlie & Robinson, 2013; Malamud & Pop-Eleches, 2011), while relatively fewer studies examine the relationship between students’ digital skills and their academic performance, generally finding a positive relationship (e.g., Hu et al., 2018; Judson, 2010; Leung & Lee, 2016; Pagani et al., 2016). However, these latter studies are correlational – being therefore susceptible to unobservable variable bias – and, in most cases, rely on self-reported measures of digital skills. Moreover, to the best of our knowledge, there is no study investigating the impact of an intervention explicitly aimed at enhancing students’ digital skills on their academic performance.
Research object and objectives
We evaluate the impact of a two-year intervention implemented in 99 lower secondary schools distributed across Italy, aimed at fostering students’ digital skills through the creation of digital communication products – such as Wiki entries, online petitions and podcasts – during Civic Education hours. Our objective is twofold: to evaluate the impact of the project on students’ digital skills and on academic performance, as measured by standardized national test scores in English, Mathematics and Italian. The project involved the presence of both treatment classes (in which the project was implemented) and control classes within each participating school. However, the assignment of classes to one of the two groups was not random but was carried out by the schools’ own teachers.
Data

The impact evaluation of participation in the project is based on two data sources: data collected as part of the project itself, through questionnaires, and data obtained from the statistical office of the Italian National Institute for Evaluation of the Educational System (INVALSI)[1].
Within the project, two questionnaires were administered to both treatment and control classes: one before the start of the project (at the beginning of the second year of lower secondary school) and one at the end of the project (at the end of the third year). These questionnaires aimed to assess through multiple-choice questions the respondents' level of digital skills, as well as to collect some background information on students' relationship with digital devices and environments.
As regards INVALSI data, we use standardized test scores for Math, Italian and English from when students were in grade 2 and 5 (second and fifth year of primary school, before participation in the project) and in grade 8 (third year of lower secondary school, upon completion of project participation).
The final sample consists of 5,986 students, including 4,619 treated and 1,367 controls, distributed throughout the Italian territory across 17 regions (29.3% in the North, 22.6% in the Centre, and 48.2% in the South).

Method

As for the empirical strategy, the first step has been to conduct balance tests on various student characteristics to assess the comparability of the treatment and control groups prior to the start of the project, given that class assignment to the treatment was not random. The balance tests were conducted with reference to student characteristics in primary school, information collected through the project pre-treatment questionnaires, and time-invariant student characteristics. Among all the variables, only the father's occupation in second grade shows a statistically significant difference between the treatment and control groups, which may reasonably be attributed to chance.
Based on this evidence, an OLS regression is adopted as the main model specification to analyse the impact of participation in the project at the individual (student) level, based on the following equation:

$$Y_{icst} = \alpha_0 + \alpha_1 T_c + \beta' X_i + \gamma_s + \lambda_t + \varepsilon_{icst}$$

where Y_{icst} denotes different outcome variables analysed in separate regressions, including INVALSI test scores in the third year of lower secondary school for Italian, Mathematics, English Listening, and English Reading, and change in digital skills score between the initial and final measurement, expressed in deciles. T_c is the treatment dummy, assuming value 1 if the student was in a treated classroom and 0 otherwise. X_i is a vector of individual level covariates, including father's occupation in second grade (found to be not balanced in the balance tests), individual participation during lower secondary school in other courses or projects related to new technologies and digital skills, allowing to disentangle the effect of participation in the evaluated project from that of similar concurrent initiatives, and INVALSI test scores in the second year of primary school for Italian and Mathematics, as a measure of baseline individual ability. School and year fixed effects are captured by the terms γ_s and λ_t, respectively. Robust standard errors are clustered at the classroom level.
To verify the robustness of the results, analyses on the impact of the project on the outcomes of interest were conducted using several alternative model specifications, including:

- The main OLS regression performed including all other control variables for which balance tests were performed;
- The main OLS regression estimated after removing potentially "always-treated" from the analysis – i.e., individuals in the control group who reported engaging in activities similar to the ones specific to the project during the project period;
- Coarsened Exact Matching (CEM);
- Inverse-probability weighting;
- Propensity Score Matching, with various matching methods (kernel, nearest neighbour with replacement, radius with a 0.0001 caliper).

As part of the robustness checks, we run some placebo tests using our main model specification and INVALSI test scores in primary school as outcome variables, to check whether selection into treatment occurred based on baseline (past) academic performance.

Results

[1] The authors are indebted to INVALSI for having allowed, preserving anonymity, the match between data from the project and the INVALSI surveys. The usual disclaimers apply.

Overall, our results indicate that the participation in the digital skills development project had a positive and statistically significant impact on digital skills (equivalent to an increase of almost 1 decile in the score distribution), while we find a negative and statistically significant effect on standardized national test scores (equivalent to 0.08 – 0.13 standard deviations from the average performance in the overall INVALSI test population), as shown in Table 1. The results for English Reading, Italian and digital skills scores are particularly robust to the alternative model specifications and matching methods. Moreover, the results of the placebo tests do not provide strong evidence of negative selection into treatment. Further analysis, through a model in which the participation in the evaluated intervention is interacted with participation in other digital-related projects during lower secondary education, suggests that the program's effect is detrimental for educational outcomes only for students who take part to additional digital activities beyond those envisioned by the program.

Table 1: Effect of participation in the digital skills development project on INVALSI test scores in 8th grade and change in digital skills score

	English Listening score Grade 8	English Reading score Grade 8	Italian score Grade 8	Math score Grade 8	Change in digital skills score
Treated	-3.262*	-4.901**	-5.326***	-3.291*	0.979***
	(1.502)	(1.581)	(1.279)	(1.416)	(0.143)
Constant	141.0***	139.7***	120.2***	103.2***	-0.152
	(4.747)	(4.787)	(4.403)	(4.931)	(0.363)
Controls	Yes	Yes	Yes	Yes	Yes
School FE	Yes	Yes	Yes	Yes	Yes
Year FE	Yes	Yes	Yes	Yes	Yes
Observations	5917	5949	5964	5951	4295
R^2	0.294	0.255	0.320	0.349	0.124

* $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

Notes: All specifications are OLS regressions. Robust standard errors clustered at the classroom level in parentheses. Controls include father's occupation in 2nd grade, student participation in other digital-related projects during lower secondary school, and INVALSI test scores in Italian and Math in 2nd grade.

Keywords: digital skills; INVALSI; academic performance; lower secondary school

SESSION 3. INTERNATIONAL LARGE-SCALE ASSESSMENTS (ILSAS) METHODS AND RESULTS

ORGANIZER: UNIVERSITY OF BATH
COORDINATOR: MARIA MAGDALENA ISAC
NOVEMBER 19TH: 2.00 P.M. – 4.15 P.M. {ROOM 5 – RESEARCH 2}

On the Compensatory hypothesis of open classroom discussion. Methodological challenges and how to address them

Diego Carrasco – David Torres Irribarra

Introduction

In the literature on civic education, the compensation hypothesis suggests that schooling can offset students' civic disadvantages (Campbell, 2008). Open classroom discussion, a recognized school effectiveness factor in civic education (Knowles, 2018), may serve this compensatory role in students' political socialization. Particularly for those from low socioeconomic backgrounds, schools provide learning experiences that may be limited or absent at home (Hoskins, Janmaat, & Melis, 2017), as open political conversations are less common in less affluent families (Bernstein, 2003). Thus, classroom discussions are expected to help counterbalance the limited political socialization of disadvantaged students (Hoskins et al., 2017).

Open classroom discussion promotes political knowledge among students (Isac, Maslowski, Creemers, & van der Werf, 2014; Persson, 2015). It characterizes classrooms where political and social issues are discussed under teachers' guidance and students are encouraged to express their opinions (Carrasco & Torres Irribarra, 2018). Moreover, it is believed to benefit all students regardless of socioeconomic status, as it does not interact with this factor (Lin, 2014), thus producing similar gains across groups.

The compensatory hypothesis is tested by including an interaction term (i.e., a product) between open classroom discussion and students' socioeconomic status in conditional models (e.g., regression, mixed models) (Campbell, 2008). If the coefficient is negative, it suggests the practice benefits disadvantaged students more than their peers, since the expected effects are smaller at higher SES and greater for lower SES students. However, prior research controlling for civic knowledge has not found a significant interaction between open classroom discussion and disadvantage (e.g., Campbell, 2008; Claes et al., 2009; Lin, 2014), concluding the practice may benefit all students similarly.

We argue that estimating interactions to assess the compensatory hypothesis faces several methodological challenges in the context of large-scale assessment studies. In the next section, we review key issues that may lead to underestimating the interaction term. Using data from the International Civic and Citizenship Education Study (ICCS), we illustrate these challenges and how they can be addressed.

Methodological challenges

Research in civic education using the International Civic and Citizenship Education Study (ICCS) (Schulz et al., 2011, 2018, 2024) face different methodological challenges when addressing research question that may require the compensatory hypothesis. In the present study we will focus on three of the most prominent problems. These are problems of model specification, model congeniality, and interacting term generation. In. the following paragraphs we will describe each problem, and how to address these.

The first is a model specification problem relative to Open Classroom Discussion scores. The Open classroom Discussion items are reference shift items, where students act as informants of a teaching practice. As such, student's collective responses from the same classroom are the referent of interest, instead of students' responses in independence of their peers (Carrasco et al., 2021). Nevertheless, previous studies have used open classroom scores in varying ways (Carrasco et al., 2018), as indicative of differences in students' experiences (see Caro and Schulz 2012; Torney-Purta 2009), and as school means to represent school attributes (see Godfrey and Grayman 2014). Yet, conditional on how these scores are included in regressions and multilevel models' users are at risk of losing their estimand of interest (Carrasco et al., 2021). In the present study, we opt for using open classroom scores aggregated at the school level, as attributed of the school learning environment.

The second problem pertains the properties of the response variable. ICCS is a large-scale assessment study (Rutkowski et al., 2010) where the scores of civic knowledge, the response variable of interest, is given to the secondary users as plausible values. Plausible Values (PV) are measurement model realizations (Wu, 2005), generated with a conditioning model (Diakow, 2013). In essence, these values can be considered a special case of imputed values. To guarantee that researchers models can retrieve an estimand of interest this estimand needs to be congenial to the imputation model use to generate the plausible values (Braun et al., 2017). Otherwise, the "non congenial estimands" may underestimates of the effect of interest. To address this second problem, we will recur to use the IRT scores (weighted least estimates, WLE) included in ICCS 2009 instead of the PV. IRT WLE scores, unlike PV, are not affected by the conditioning model. Although, IRT WLE could potentially lead to bias due to not accounting for measurement error, ending in attenuated estimates (Bhaktha et al, 2021), these biases are ameliorated when the test includes a large number of items (e.g., more than 25 items), has high reliability, and a spread distribution of item locations (Diakow, 2013).
When we students nested in schools/classrooms and the factor of interest are at different levels, there are different alternatives regarding how to **create interacting terms**. In the present case we have a student attribute, an indicator of student disadvantage, and school attribute, a teaching practice factor, generated with students' responses, the scores of Open Classroom Discussion. Ryu (2015) simulation studies shown that if the level 1 variable is interacted with a level 2 factor, the obtained results can be biased conditional to the type of centering used for the variables (e.g., uncentered, centered within cluster, centered to the grand mean). In this study, we create the interacting term of interest between the disadvantage factor, centered to the grand mean with the cluster means of Open Classroom discussion; to assure we can retrieve the estimate of interest.

The present study.

To illustrate the previous problems, we will fit a series of regression models using data from the International Civic and Citizenship Education Study (ICCS 2009). We will address the following research questions:

a) Can we identify the same countries as presenting significant interacting terms between disadvantage factors and open classroom discussion, while using plausible and while using IRT WLE scores?

For this line inquiry we hypothesize that PV should underestimate the interacting effects of interest due to the non-congeniality of the interacting term. In contrast, we expect that models conditioning on IRT WLE scores may include more countries with the effect of interest.

b) Can we find similar compensatory effects regardless of the disadvantage factor?

Here we are expecting to find similar results for socioeconomic status, parents' educations and, number of books at home. Because socioeconomic status is composite score that includes parents' educations, and the number of books at home, we expect to find convergent results across all these sent of indicators.
In the following section, we describe the methods and models used to illustrate the reviewed problems.

Methods.

Data.

We used data from the International Civic and Citizenship Education Study from 2009 (ICCS 2009). This study provides representative representative samples of 8th grade students, using a two-stage design, with intact classrooms. To illustrate the present problems, when estimating interactions with plausible values, we fit regression models (while accounting for the study sampling design). From the present study, we exclude Luxemburg, Malta, and Liechtenstein due to low count of schools ($n < 60$). As such, we are including in this exercise the 35 remaining educational systems. ICCS 2009 is a convenient ILSA study, because it provides with PV and IRT WLE estimates from the same test (Schulz et al., 2011).

Dependent variable. Civic Knowledge (y_ij) scores represent students' political understanding of political issues. It consists of five plausible values, generated with IRT Rasch model over a random booklet design of 79 item-test. We will use the provided plausible values (PV1-PV5), and the IRT WLE scores from each country generated with the 79 adjudicated items for the international test. Each of these scores were standardized using the international pooled sample mean and standard deviation for each.

Disadvantage factor (Independent variable). To illustrate the presented problems, we are including socioeconomic Status (ses_ij) of student's families, parents' education (edu_ij) and number of books at home (bok_ij). Socioeconomic status is composite scores based on the responses to the items addressing student Parents Education level, Parents Occupation, and number of books at home. This is a continuous variable, that we have standardized at the country level. Additionally, we created dichotomous version of parents' education (edu_ij) and number of books at home (bok_ij), to distinguished disadvantage students across

countries (parents with non-tertiary education, and less than 100 books at home respectively), from students from more afluent homes (students with at least one parent with a teriary education degree; and students from homes with more than 100 books, respectively).

School factor (Moderator variable). Open classroom discussion scores are partial credit model realizations with a mean of 50, and standard deviation of 10 values for the international pooled sample. In the present study for illustrator purposes, we have standardized the original scores, so these have a mean of 0, and standard deviation of 1.

Analytical strategy

To illustrate the problem of estimating interactions to assess the compensatory hypothesis we fit the following model, onto each country sample:

$$y_{ij} = \alpha + \beta_1(x_{ij} - \bar{x}_{..}) + \beta_2(w_{.j} - \bar{w}_{..}) + \beta_3(x_{ij} - \bar{x}_{..})(w_{.j} - \bar{w}_{..}) + \epsilon_{ij}$$

In the present notation we use x_{ij} to represent our disadvantage factors (i.e., ses_ij, edu_ij, bok_ij), and use $w_{.j}$ to referred to the scores of open classroom discussion. Our estimand of interest is β_3, which is the coefficient associated with interaction term. If this estimate is different from zero and negative, we would have evidence favoring the compensatory hypothesis. Conversely, if β_3 is different from zero and positive our results would favor the acceleration hypothesis (Campbell, 2008), where more advantage students benefit more than their peers from open classroom discussion. In consequence, we fit 210 models, by varying 2 response variables, 3 covariates, and 35 different educational systems, participating in ICCS 2009. Our estimand of interest is β_3. All estimates are pseudo maximum likelihood estimates, where survey design weights and plausible values are accommodated accordingly (Rutkowski et al., 2010), when pertinent.

Preliminary results.

Using the methods described above, we found significant interacting terms in higher proportion when the response variable was the IRT WLE scores (21%), when we used the PV (17%). In all instances the interaction term was positive, implying an accelerating effect instead of compensatory. That is, that students from more advantage backgrounds benefit more from Open classroom discussion, than students from disadvantage backgrounds, with one exemption. We found a negative interacting term when the dis-advantage factor was parents' education in Chile. Regarding our second line inquiry, we found that although models with IRT-WLE scores have overall higher chances of finding $\beta_3 \neq 0$ in contrast to models using plausible values, this differences was larger when the disadvantage factor was socioeconomic status (ses_ij) (17% with PVS, 26% with IRT-WLE), than when the disadvantage factor was Parents education (edu_ij), or the number of books at home (bok_ij).

Conclusion and Discussion.

In the present study, we have shown that the study of interacting terms to assess the compensatory hypothesis of open classroom discussion from PV and IRT WLE scores do not lead to the same conclusions. PV imputations do not recover interacting terms due to its non-congeniality. Its alternative, models fitted including IRT WLE scores are better alternative for the present purpose. Further research is needed to identify when PV and IRT-WLE disagree the most.

The present exercise has two implications. The first pertain secondary users, and the second pertains ILSA study operations. Secondary researchers need to address the limitations of the methods they are using when making inferences. This is specially the case if the chosen methods obstacle their inferences of interest. On the other hand, ILSA studies operations, aiming to promote the study of educational gaps with risk and protective factors at the school level are encouraged to vary their PV method generation accordingly, and provide thorough documentation of how these are generated. Thus, allowing secondary users to identify the limitations of the generated scores. ICCS studies, are good example of how thorough this documentation can be (see Schulz et al., 2011, 2018, 2022 for examples).

Keywords: Civic education, open classroom discussion, compensatory hypothesis, plausible values

Who Welcomes Whom? Understanding Socio-Economic Disparities in Students' Support for Equal Rights for Immigrants

Francisco Gatica – Jan-Philip Wagner

Human mobility is a dynamic and growing global phenomenon. As of 2023, over 117 million people were forcibly displaced worldwide, according to the United Nations Refugee Agency. The causes of migration range from voluntary and temporary relocation to asylum-seeking and forced displacement. These experiences intersect with factors such as gender, socioeconomic status, and religion, shaping immigrants' opportunities for social inclusion. In this context, migration challenges education systems to rethink how civic and citizenship education is delivered in a globalized world, guided by principles of tolerance and equality.IEA'S International Civic and Citizenship Education Study among lower-secondary students provides valuable information about educational systems, allowing research and cross-country comparisons. ICCS investigates how young people are prepared to undertake their roles as citizens in a world where contexts of democracy and civic participation continue to evolve. Equity as a subdomain in the ICCS 2022 framework focuses on the principle that all people have the right to fair and just treatment, therefore protecting and promoting equity is essential to achieving peace, harmony, and productivity within and among communities (Schulz, Fraillon, et al., 2023).

According to the most recent 2022 cycle results, lower-secondary students strongly support for equal rights for immigrants. Nevertheless, results also showed differences associated with socioeconomic background and civic knowledge (Schulz, Ainley, et al., 2023). Studies have shown that positive attitudes towards equal rights are associated with both individual and school-level factors. Background characteristics such as parental education and the number of books at home are predict more inclusive attitudes (Miranda et al., 2018), along with high levels of civic knowledge, socioeconomic status, gender, and students' high expectations about their future and interest in social and political issues (Isac & Claes, 2020; Schulz & Ainley, 2018; Isac et al., 2012). At the school level, Isac and Miranda (2025) note that an open classroom climate is positively associated with students' support for equal rights for immigrants. However, studies using ICCS data suggest that school-level factors account for only a small variation in students' attitudes toward immigrants' rights (Blaskó et al., 2019; Isac et al., 2018). Despite efforts to understand students' attitudes toward immigrants within the context of civic and citizenship education, few studies using large-scale assessment data have examined whether the influence of individual- and school-level factors on these attitudes varies across socioeconomic groups. To address this gap, the present study applies an Oaxaca-Blinder decomposition approach to investigate the extent to which differences in students' positive attitudes toward immigrants between different socioeconomic groups can be accounted for by these factors. This aligns with school effectiveness research on differentiated effectiveness, which explores how the same explanatory factors affect student outcomes differently (see Reynolds et al., 2014).

Method

To identify which factors explain the gap in students' support for equal rights for immigrants across socioeconomic groups, this study applies the Oaxaca-Blinder decomposition (OBD) method. Originally developed to analyse wage gaps (Blinder, 1973; Oaxaca, 1973), this technique has been widely used in the field of education to examine disparities in learning outcomes across dimensions such as gender (Gevrek et al., 2020; Fortin et al., 2015), school sector (Parvez & Laxminarayana, 2022; Romuald, 2023; Castro Aristizábal et al., 2017), Indigenous student status (Arteaga & Glewwe, 2019; Blanco, 2019).Following the linear model formulation proposed by Jann (2008), the OBD estimates how much of the observed difference between two groups—in this case, students in the top and bottom quartiles of socioeconomic status—can be attributed to group differences in predictors. The total difference is decomposed into three components: endowment effect (differences in characteristics), coefficient effect (differences in returns to characteristics), and interaction effect (simultaneous differences in both). In this study, the decomposition is estimated from the perspective of the low socioeconomic group: the endowment effect reflects the expected change in attitudes if low-SES students had the same predictor values as high-SES students, while

the coefficient effect represents the expected change if low-SES students had the same regression coefficients as their high-SES peers.Following the ICCS design, appropriate student, class, and school weights were applied for multilevel analysis (Schulz et al., 2024).The OBD is estimated separately for each participating country. Analysis was conducted using WeMix package in R.

Data

The study uses data from ICCS 2022 across all participating countries. The outcome is students' positive attitudes toward immigrants (S_IMMPOS).

Individual-level predictors:

- Immigrant background (IMMIG_B): 0 = no; 1 = yes
- Gender (SD_GENDER): 0 = boy; 1 = girl
- Civic knowledge (PV1CIV)
- Perceived openness in class discussions (S_OPDISC)

School-level predictors:

- Proportion of immigrant students (IMMI_SCHOOL)
- Classroom mean civic knowledge (CIVIC_SCHOOL)

Preliminary results

In most countries, the gap in students' support for equal rights for immigrants between the bottom 25% and top 25% of the socioeconomic status distribution ranges from 2 to 3 points on the S_IMMPOS scale, corresponding to approximately 0.2 to 0.3 standard deviations.

Civic knowledge

The endowment effect for civic knowledge varies between 2.5 and 3 in most countries, indicating that high-SES students tend to score 2.5 to 3 points higher on the S_IMMPOS scale due to their higher levels of civic knowledge. In other words, if low-SES students had the same level of civic knowledge as high-SES students, their scores would be expected to increase by this amount.

However, in most countries, we observed a negative value for the civic knowledge coefficient, indicating that civic knowledge is less strongly associated with positive attitudes among low-SES students. That is, even if low-SES students had the same level of civic knowledge, the "return" of that knowledge—in terms of fostering positive attitudes—is lower. This suggests that civic knowledge at an individual level is *less effective* in promoting positive attitudes toward immigrants among low-SES students.

Classroom Mean of Civic Knowledge Outcomes

The endowment effect of the classroom mean for civic knowledge is relatively small (around 0.2) across most countries. This suggests that low-SES students would not experience a substantial increase in their positive attitudes towards immigrants even if they were in classrooms with the same average level of civic knowledge as their high-SES peers.

In contrast, the coefficient effect tells a different story. In ten countries this effect ranges from 4 to 10 points, indicating that the influence of classroom-level civic knowledge on students' attitudes is significantly stronger among low-SES students. In these contexts, being in a classroom with higher civic knowledge appears to be especially beneficial for students from disadvantaged backgrounds, potentially amplifying their support for immigrants' rights.

Immigrant background and classroom composition

The endowment and coefficient effects for immigrant background (i.e., whether the student or their parents were born outside the country of the test), and classroom immigrant composition are relatively small across countries. However, despite the generally small effects, in eleven countries low-SES students tend to show more positive attitudes when exposed to a higher share of immigrant peers.

Students' perceptions of openness in classroom discussions

No consistent pattern emerged for perceptions of openness in classroom discussions. The endowment effect, although relatively small in all countries, is positive in the majority. Nonetheless, the coefficient effect varies in both magnitude and direction. In five countries, the coefficient is positive and above 1 suggesting that a greater perception of openness in classroom discussions could be effective. On the other hand, in others (seven countries) the coefficient is negative suggesting that a greater perception of openness in classroom discussions is less effective.

Conclusion

This study reveals important differences between low and high socioeconomic status (SES) students in their positive attitudes toward immigrants, providing new insights into how individual and school factors operate differently across these groups using OB decomposition. Consistent with existing literature, civic knowledge plays a crucial role in fostering positive attitudes toward immigrants. However, our preliminary findings also indicate that the same level of civic knowledge tends to be less effective in promoting these attitudes among low-SES students compared to their high-SES peers. This suggests that simply improving civic knowledge may not be sufficient to close attitudinal gaps without addressing the broader social and contextual barriers faced by disadvantaged students.

Moreover, the classroom environment matters, particularly the average level of civic knowledge within classrooms. In some countries, low-SES students benefit significantly when surrounded by peers with higher civic knowledge, highlighting the potential of classroom composition to amplify positive attitudes.

Finally, perceptions of openness in classroom discussions show mixed effects across countries, underscoring the importance of considering national and cultural contexts when interpreting these results.

Keywords: Civic Attitudes, Inequality, Immigrant Rights, Oaxaca-Blinder Decomposition

Gender Attitudes among Adolescents: A Systematic Review of International Quantitative Studies (2000–2024)

Natalia López-Hornickel – Belen Gutiérrez de Rozas – Jesús Vidal – Andres Sandoval-Hernández

Gender equity attitudes are key components of democratic citizenship (Schulz et al., 2018). According to Social Identity Theory (Hogg & Smith, 2007) and Political Socialisation Theory (Abendschön, 2013), these attitudes are shaped during adolescence through family, school, and broader socio-cultural influences.

This systematic review investigates how quantitative studies between 2000 and 2024 have examined adolescents' gender attitudes, focusing on research trends, geographical scope, study aims, theoretical underpinnings, and methodological strategies.

Although this review does not use INVALSI data directly, it offers relevant insights for future exploitation of ILSA datasets (e.g. ICCS, PISA) by highlighting the methodological potential of analysing gender and civic attitudes in the Italian context.

Following PRISMA guidelines (Moher et al., 2009), 137 peer-reviewed articles were selected. The results show a marked increase in publications after 2015, peaking in 2022. Only 32% of studies explicitly apply theoretical frameworks, and fewer than 10% examine school-related factors. Most research focuses on intergenerational transmission (24%) and intervention evaluations (13%), while studies exploring school influence, intersectional perspectives, or civic education remain limited (e.g. Dotti Sani & Quaranta, 2017).

Geographically, studies are concentrated in Asia (29%), North America (USA) (26%), and Europe (22%). Latin America and Africa are notably underrepresented, and southern European contexts such as Italy receive minimal attention. This reflects a gap in understanding how gender attitudes develop in diverse educational systems, particularly in under-researched regions (López-Hornickel & Sandoval-Hernández, 2023).

In terms of scale measurement, it is interesting to observe that most studies have developed their own data collection (primary data) in their contexts, adapting instruments such as the Gender-Equitable Men (GEM) scale (Pulerwitz & Barker, 2008) to their settings, which usually implies the use of Likert scales to evaluate the agreement with gender attitudes statements. This brings interesting discussions regarding external validity, since not all data has a random design in their collection. While this raises questions about

generalisability, such approaches address an urgent need to understand gender attitudes in contexts often excluded from large-scale comparative assessments (Scott, 2010; D'Ignazio & Klein, 2023).
These findings reveal clear trends and omissions in the current literature. The review provides methodological insight into how gender attitudes are operationalised in ILSA-type research and highlights the need for more theory-informed, education-relevant analyses. It encourages the use of existing large-scale datasets to examine underexplored non-cognitive educational outcomes, including gender attitudes and democratic values, especially within the Italian system.

Keywords: Quantitative approach, Gender Equity Attitudes, Adolescence

Gamson's Mistrustful-Efficacious Political Participation Hypothesis in adolescents from 24 countries using the ICCS 2022 study

Daniel Miranda – Felipe Sánchez-Barría

Young people have played a prominent role in the global trend towards protest participation to the detriment of more traditional forms of political involvement. In recent years younger generations have not only been the drivers and protagonists of the protest but often exceeded their more peaceful and "normalized" forms to make their demands heard. What drives the youngest to get involved in these highly disruptive actions? What differentiates them from those who opt for other moderate avenues of political action? How and to what extent can the interaction between individual characteristics and context radicalize young people?
William Gamson (1968) proposed that low institutional trust and high internal political efficacy are optimal for political mobilization, including participation in protests and demonstrations. According to Gamson, mobilization against the government occurred primarily because of dissatisfaction with its performance but, at the same time, because they believed they could bring about social change through direct action. This paper follows this hypothesis, using the combination of institutional trust and political self-efficacy to generate typologies of "political participants" and assess the extent to which the expected pattern occurs considering different types of participation: conventional, disruptive and digital.
Using the International Civic and Citizenship Education Study 2022 - ICCS, around 80000 students from 24 countries participated (Schulz, 2025), we evaluated Gamson's Mistrustful-Efficacious Political Participation Hypothesis. The key independent variable is the "Gamson typology", created by combining high and low levels of self-efficacy with high and low levels of institutional trust. To achieve this, we use the scale Students' citizenship self-efficacy (S_ CITEFF) and institutional trust variables. Students' citizenship self-efficacy is an IRT WLE scores with mean of 50 and standard deviation of 10 across equally weighted participating countries that met sample participation requirements, based on seven items regarding the degree of agreement with the following statements: IS4G29A: Argue your point of view about a controversial political or social issue, IS4G29B: Stand as a candidate in a [school election], IS4G29C: Organise a group of students in order to achieve changes at school, IS4G29D: Follow a debate about a controversial issue, IS4G29E: Write a letter or email to a newspaper giving your view on a current issue, IS4G29F: Speak in front of your class about a social or political issue, IS4G29G: Assess the credibility of information about political or social issues. The Students' trust in civic institutions (S_INTRUST) is an IRT WLE scores with mean of 50 and standard deviation of 10 across equally weighted participating countries that met sample participation requirements, based on six items assessing the degree of trust that students have in the following institutions: IS4G27A: The [national government] of [country of test], IS4G27B:The [local government] of your town or city, IS4G27C: Courts of justice, IS4G27D: The police, IS4G27E:Political parties, and IS4G27F: [Parliament/congress].

Once we have obtained self-efficacy and political confidence measures, we construct the "Gamson typology," consisting of four groups. To create this typology, we defined low and high levels of self-efficacy and political confidence using the 25% and 75% quartiles as cutoff points to categorize students within each group. Thus, those youth who scored above the 75% quartile on internal efficacy and below the 25% quartile on political confidence were categorized as "Dissidents." While respondents who scored above the 75% quartile on the efficacy and trust variables were labeled the "Assureds." On the other hand, the "Alienated" scored below the 25% quartile on self-efficacy and institutional trust. We classified as "Subordinates" those students who combined high trust in institutions, above the 75% quartile but were

below the 25% quartile in terms of their self-reported efficacy level. Using quartiles as high and low cutoff points, rather than the mean, allows for a more apparent distinction between 'average' individuals and those with relatively more extreme attitudes or political orientations (Zhang & Lin, 2018). Finally, those youth who reported intermediate levels of self-efficacy and institutional trust are classified as "Moderates." While this category falls outside the "Gamson typology," it represents the sample's more "typical"(average) individuals.

In this study, we use three dependent variables that measure different forms of participation. The first, *students' expected participation in legal activities* (S_LEGACT), composed of various items that ask whether the students has the intention to participate in different actions: IS4G31A: Talk to others about your views on political or social issues, IS4G31B: Contact an [elected representative], IS4G31C: Take part in a peaceful march or rally, IS4G31D: Collect signatures for a petition, IS4G31E: Contribute to an online discussion about social or political issues and IS4G31F: Organise an online campaign in support of a political or social issue. The second variable is *students' expected participation in illegal protest activities* (S_ILLACT), composed of various items that ask whether the students has the intention to participate in different actions: IS4G31K Spray-paint protest slogans on walls, IS4G31L: Stage a protest by blocking traffic, and IS4G31M: Occupy public buildings as a sign of protest. The third variable is *students' engagement with political or social issues using digital media* (S_ENGDM), composed of various items that ask whether the students' involvement in different actions: IS4G13A: Using the internet to find information about political or social issues, IS4G13B: Posting your own content about a political or social issue on the internet or social media, IS4G13C: Sharing content about a political or social issue posted by someone else, IS4G13D: Commenting on an online post about a political or social issue, and IS4G13E: [Liking] an online post about a political or social issue.

We estimate a series of linear regression models to evaluate the hypothesis to assess the relationship between Gamson´s typology and political participation types. In all analyses, the reference category were "Moderates." Any significant difference observed in each category within the "Gamson typology" should thus be interpreted as a difference concerning the more typical group. The absence of a statistically significant association would imply that this category is not different from the "average" group concerning each specific type of participation.

The results show that the typology most willing to participate in legal activities is assured (efficacious and trusted) and then dissidents (efficacious and distrusted) in all countries, although differences are observed in the predicted levels for each country. Regarding illegal participation, as expected according to Gamson's hypothesis, those most willing to participate are dissidents, and then assured. However, in some countries, this order of priority is reversed, with assured students showing a greater willingness to participate in illegal activities. This demonstrates the relevance of self-efficacy as an explanatory factor for participation. Something similar to the pattern of unlawful participation occurs with digital participation. Assured and dissident students tend to show the highest levels of involvement using digital media.

In line with expectations, results showed that efficacious students (trusting and distrustful) are more willing to participate in the different types. These analyses are still in development, but they help demonstrate the relevance of the old-fashioned Gamson hypothesis. Although Gamson theorized about this during the 1960s, it still reflects its relevance in accounting for the changing and new forms of participation among the youth population. For the presentation, we will further compare countries and the regional differences that can be observed.

Keywords: Gamson's Mistrustful-Efficacious Hypothesis, Digital participation, disruptive participation, conventional participation.

Beyond the Bifactor: Identifying Person Heterogeneity Effects in the ICCS Gender Equality Attitude Scale

David Torres Irribarra – Diego Carrasco – Fernando Ponce

Introduction

Understanding students' attitudes toward gender equality is a central concern in civic and citizenship education research. The International Civic and Citizenship Education Study (ICCS) (Schulz et al., 2011, 2018, 2024) includes a six-item scale (GENEQL) designed to measure support for gender-equal rights and roles, encompassing domains such as political participation, employment equity, and leadership. Prior research has supported both unidimensional and two-factor solutions for this scale (e.g., Castillo, Miranda & Bonilla, 2021), often distinguishing between positively and negatively worded items. However, such models frequently conflate content with method variance, without adequately addressing the complexity introduced by item wording.

Bifactor models are commonly used to address these concerns by partitioning variance into a substantive general factor and an orthogonal method factor. Yet recent methodological work (Ponce et al., 2022; Ponce, Torres Irribarra, Vergés, & Arias, 2023) suggests that these solutions may obscure critical respondent-level heterogeneity. Specifically, they demonstrate that wording effects are not uniformly distributed across populations, but instead emerge from specific subgroups whose response patterns diverge meaningfully from the general trend.

In this study, we apply a person-specific method approach to reanalyze the dimensional structure of the GENEQL scale. Rather than assuming that item wording effects generate multidimensionality at the population level, we combine latent class modeling with confirmatory factor analysis to detect which subgroups of respondents are susceptible to these effects. This strategy enables us to determine whether more parsimonious measurement models can appropriately describe a substantial proportion of the population, avoiding the unnecessary complexity of imposing bifactor or multidimensional models across the board.

Importantly, we extend prior applications of this methodology by incorporating ICCS's complex sampling design—including stratification, clustering, and sampling weights—into all phases of the analysis. Thanks to the use probabilistic samples from different countries, the present results may support more extensive generalizability across contexts. Although previous studies have explored the response patterns onto GENEQL items as mixtures (see Inostroza, 2021; López-Hornickel et al., 2023), the present study does so with specific focus of identifying person-specific method effects. All in all, by integrating these elements, our analysis offers a more robust evaluation of the factorial structure of gender equality endorsement measures in large-scale international assessments.

Research Objectives and Hypothesis

This study investigates the dimensional structure of the GENEQL scale in ICCS 2009, which measures students' support for gender equality. Prior models (e.g., bifactor or two-factor structures) often attribute multidimensionality to item wording effects, treating them as population-wide phenomena. However, recent methodological insights suggest that such effects may be concentrated in identifiable respondent subgroups.

Our hypothesis is that method variance is not uniformly distributed across the population, but instead reflects person-specific response patterns. We expect to find that a more parsimonious unidimensional factor structure is sufficient for the majority of students once those affected by strong wording-related effects are identified and excluded.

Methodological Framework

Data

We use data from ICCS 2009, comprising responses from over 330,000 students across 38 countries. The six-item GENEQL scale includes three positively worded items and three reverse-coded items that address gender equality in political participation, employment, and leadership.

In addition to attitudinal responses, the dataset includes a range oreview it f relevant covariates: Age, Gender, Socioeconomic status, Expected educational attainment, and Parental education. These covariates allow us to explore how response patterns and subgroup membership relate to key background characteristics.

We account for ICCS's *complex sample design*, including stratification, clustering, and sampling weights, throughout all phases of the analysis via using Pseudomaximum Likelihood and Taylor Series Linearization (Asparouhov, 2005; Stapleton, 2008).

Methods

We use data from ICCS 2009, comprising responses from over 300,000 students across multiple countries. The GENEQL scale includes six items (three positively and three negatively worded) related to gender roles in civic life. Covariates include gender, age, socioeconomic status (SES), and parental education.

We first apply Latent Class Analysis (LCA) to identify subgroups based on response patterns, particularly to detect potential method-related heterogeneity.

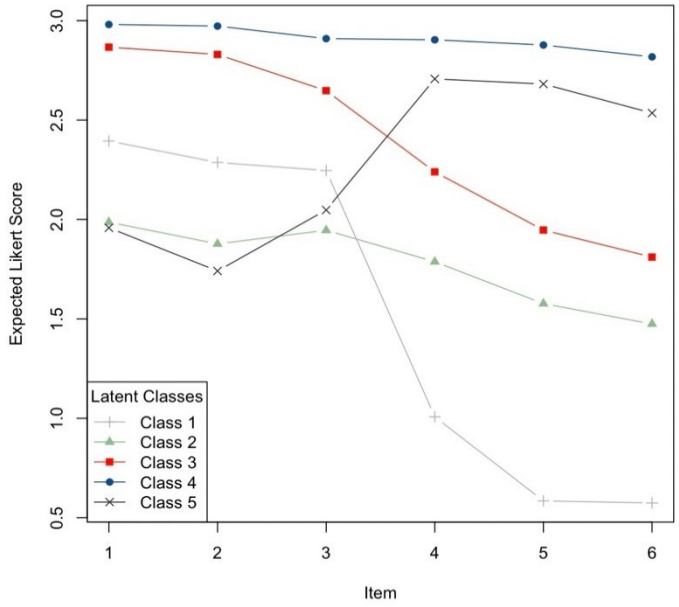
We then estimate Generalized CFA models using WLSMV estimation, comparing model fit for the full sample and for a depurated subsample excluding the identified method classes. Model fit is assessed using CFI, TLI, SRMR, and RMSEA. Future analyses will integrate latent class predictors and explore ordinal-specific models to account for imbalanced category selection and further refine the measurement structure.

Results and Findings

Preliminary Results

Latent Class Analysis and Subgroup Identification. We conducted a latent class analysis (LCA) of the six-item GENEQL scale, identifying five distinct response profiles (see Figure 1). These classes illustrate heterogeneity in how students interpret and respond to positively worded versus reverse-coded items.

Figure 1. Expected Item Scores by Class (5-Class LCA)



- **Class 1** (10% of the sample) displayed high endorsement of positively worded items and low endorsement of reversed items—consistent with method-sensitive response patterns described by Ponce et al. (2022, 2023).

- **Class 5** (4%) exhibited the *inverse* pattern, favoring reversed items over direct ones.

Together, these two classes account for **14% of respondents**, indicating that item wording effects are not uniformly distributed but instead **concentrated in distinct subgroups** of students. The remaining 86% of the sample showed relatively stable and interpretable ordered patterns across both item types.

Generalized CFA and Model Fit. To assess the impact of these subgroups on model structure, we estimated a **unidimensional generalized CFA** model on both the full dataset and a depurated subsample excluding Classes 1 and 5.

In the full sample (n = 315,342), model fit was modest to poor across several indices:

- CFI = 0.931, TLI = 0.885, SRMR = 0.079
- RMSEA = 0.156 (90% CI: 0.155–0.157)

In the refined sample (n = 270,763), excluding Classes 1 and 5 (14% of respondents), fit improved across all indices:

- CFI = 0.984, TLI = 0.974, SRMR = 0.025
- RMSEA = 0.101 (90% CI: 0.100–0.102)

These results show a substantial and consistent improvement in model fit once subgroups driving method effects are removed. While CFI, TLI, and SRMR reach levels typically considered acceptable (e.g., Hu & Bentler, 1999; Rutkowski et al., 2017), RMSEA, while still elevated, showed marked improvement and may reflect distributional issues. This may be due to imbalanced response category use, non-normality, or limitations of the unidimensional model itself. However, RMSEA results needs to be interpreted with caution because its ability to detect correctly specified models with a single cut off value is not guaranteed (e.g., Maydeu-Olivares et al, 2018), while SRMR is a more robust alternative (Shi et al, 2020).

Planned Extensions

Our presentation will include a full comparison of alternative models to determine whether simpler or more flexible models offer better explanatory power:

A **two-factor model** separating positively and negatively worded items,

A **bifactor model** that isolates a general factor and an orthogonal method factor,

A **person-specific method model** that integrates latent class membership into the measurement structure.

We will evaluate each model using standard fit indices (CFI, TLI, RMSEA, SRMR) and class separation statistics. Additionally, we plan to consider ordinal-specific models and alternative link functions to account for imbalanced category usage and non-normal response distributions observed in preliminary analyses.

Additionally, we will conduct a focused analysis of the Italian sample from ICCS 2009 to examine whether the observed response profiles and method effects replicate within this national context, thereby enabling targeted insights into the measurement of civic attitudes in Italy.

To further refine the measurement structure and clarify the role of person-specific method effects, we will also extend our analyses by predicting latent class membership using key covariates, including student gender, age, socioeconomic status, and parental education, to understand which population segments are more likely to exhibit method-driven response patterns.

Implications

Our findings have important implications for how reversed items are handled in attitude measurement, especially in international large-scale assessments. By identifying when and for whom method effects distort item responses, we provide an alternative to overfitting multidimensional models. This person-specific approach enhances construct validity, clarifies interpretation, and offers methodological guidance for future instrument design and validation.

By zooming into the Italian sample, we aim to assess the extent to which these method-related patterns apply in the national context. This will help strengthen the validity of civic and citizenship attitude measures used in ICCS and similar studies in Italy. More broadly, the findings contribute to methodological innovation within ILSAs and provide empirical support for improving cross-national comparability and policy relevance.

Conclusion

This study presents a person-specific reanalysis of the GENEQL scale in ICCS 2009, illustrating how latent class modeling can reveal the **localized nature of method effects**. By combining this with confirmatory factor models and incorporating ICCS's complex sampling design, we aim to refine the dimensional interpretation of gender equality attitudes. Our findings point to more robust and interpretable models of student belief systems, with implications for both psychometric practice and international policy evaluation.

Keywords: International Large-Scale Assessments, Method Effects, Latent Class Analysis, Civic Attitudes

Gender differences in perceptions of open classroom discussion using ICCS 2022

Ernesto Treviño – Catalina Miranda – Diego Carrasco

Introduction

Open classroom discussion is a key predictor of civic and citizenship outcomes among 13-year-old students around the world (Carrasco et al., 2020; Treviño et al., 2017, 2021). Schools regarded by students as places with positive climates to openly discuss different perspectives show higher dispositions to participate in formal politics through the political systems, as well as to participate in social movements at the community level (Torres Iribarra & Carrasco, 2021; Treviño et al., 2017, 2018, 2021). Therefore, classroom discussion is a key process to develop citizenship skills among the youth.

Due to the importance of open classroom discussion as a predictor of citizenship attributes, it is necessary to delve into the different ways in which female and male students within the classroom perceive the openness of classroom discussion. In educational research, there is plenty of evidence of how the participation of girls and boys within the classroom may be unequal (Black & Radovic, 2018; Meer & Lim, 2017; Minasyan, 2017; Ortega et al., 2021). The perceptions on classroom climate may be related to differential disruptive behaviors between females and males (Van Houtte, 2024); moral motivations for civic engagement (Malin et al., 2015); and political efficacy (García-Albacete & Hoskins, 2025), among others.

Following this idea, this paper studies the differences in open classroom discussion perceptions among girls and boys in the countries participating in ICCS 2022, following a similar logic to previous work, which did not investigate gender differences (Carrasco et al., 2021). It is important to note that, in other civic attributes different from open classroom discussion male students hold consistent advantages in comparison to female students. This proposal aims to delve into the gender gap in open classroom discussions and the factors that may explain such differences.

Research hypotheses

The purpose of this study is to investigate the differences in perceptions of open classroom discussion among female and male students who participated in the International Civic and Citizenship Education Study (ICCS) 2022, as well as the variables that may explain such gaps.

The main hypothesis of the study is that there are differences in the perception of open classroom discussion between female and male students, a trend that varies by country, and is explained by different school and contextual variables. The main explanatory variables are the proportion of female students in the school, the type of school (public or private), the size of the school; students' participation in the school achievement in civic knowledge; and expectations of future participation are the main predictors used in this study.

Data

The study uses the data from the ICCS 2022, which collects information from 13.5-year-old students in 22 countries and two regions of Germany. The sample of the international study comprised nearly 3,500 schools, 50,000 teachers, and 80,000 students. This means that principals, teachers, and students completed a series of questionnaires to collect information on civic and citizenship attributes such as students' civic knowledge, attitudes, engagement, as well as contextual factors at home and the characteristics of schools and teaching.

Method

The study uses generalized linear models in the form of multilevel regressions. This methodology is necessary to recognize that students are nested within schools and, therefore, correct for the lack of independence of those attending the same school (Rabe-Hesketh et al., 2012; Raudenbush & Bryk, 2002; Snijders & Bosker, 2012). The method implies, first, estimating the intraclass correlation of the open classroom discussion index. Secondly, the method allows for testing if there are statistically significant differences in the perception of open classroom discussion within and between schools. In third place, the method permits the modeling of factors (type of school, socioeconomic status, ethnic background, proportion of females in the school, among others) that explain the gap in perceptions of open classroom discussion between females and males school (Rabe-Hesketh et al., 2012; Raudenbush & Bryk, 2002; Snijders & Bosker, 2012).

The study fits a series of multilevel models to test the hypothesis. First, it analyzes the results for the complete database of 22 countries and two regions to estimate the gender gap in open classroom discussion. Second, it fits a series of multilevel models introducing country fixed-effects, school, and student variables

as predictors of the gender gap. Third, it fits a series of models separately for each country as a way of robustness check.

Results and Findings

Preliminary results show that there are gender differences in the perceptions of open classroom discussion, although it portrays advantages for females instead of males. Such a trend is general among the participating countries, although there are exceptions to this pattern. Therefore, it seems that female students seem to have more opportunities for open classroom discussion than males. This may be related to the identification of female students with the school and classroom cultures, cultivating positive relationships with their teachers (Bugler et al., 2016; Kang et al., 2023).

These gender differences are explained by three different factors. First, it seems that country cultures explain a portion of the differences found among countries. Second, school characteristics such as the proportion of female students in the school, the type of school (public or private), and the size of the school are the main predictors of the gaps. Third, at the student level, their participation in the school, their achievement in civic knowledge, and their expectations of future participation predict the gap in open classroom discussion. However, the weight of the explanation varies by country.

Keywords: Gender, school participation, civic education, citizenship

A Systematic Literature Review: Categorizing and Interpreting Criticism of the International Civic and Citizenship Education Study (ICCS)

Jan-Philipp Wagner – Francisco Gatica

The International Civic and Citizenship Education Study (ICCS) commissioned by the International Association for the Evaluation of Educational Assessment (IEA) assesses the ways in which different education systems prepare their students for their future roles as citizens. Together with its precursor, the Civic Education Study (CIVED), ICCS has had considerable amounts of influence on practice, policy and research in different education systems and regions (Malak-Minkiewicz & Torney-Purta, 2021). Several literature reviews have demonstrated the continued impact of data from ICCS and CIVED on educational research, large numbers of scholars from a variety of national and disciplinary backgrounds undertaking secondary analyses into a wide range of research questions (Knowles & Di Stefano, 2015; Knowles et al., 2018).

However, there is no literature review of the criticism that ICCS (and CIVED) has received since its first study cycle in 2009. In an attempt to fill this gap in the literature, this paper offers a systematic literature review of publications that are critical of ICCS. To this end, the literature review is guided by two primary research questions: (1) How can critiques of ICCS be categorized into broader strands of criticism? (2) How can these strands of criticism be understood in the context of the characteristics of international large-scale assessments (ILSA), with the aim to identify actionable improvements to ICCS? The findings of the categorisation of criticism and its interpretation can form the basis for targeted improvements to ICCS and facilitate responses to the critical voices, thereby furthering scholarly debate among the research community.

The primary methods used in this paper follow the guidelines of the Preferred Reporting Items for Systematic reviews and Meta-Analyses (PRISMA) 2020 statement (Page et al., 2021). To narrow the search, the literature review does not include critiques of ILSAs in education in general but focusses specifically on critiques of the ICCS-precursor CIVED (1999), ICCS 2009, ICCS 2016, and ICCS 2022.

Preliminary results suggest that published critiques of ICCS (and CIVED) can be categorized in three groups. First, normative (or value-based) criticism focuses on the values and normative considerations that underpin ICCS, according to the critics. For instance, some scholars argue that the content and design of ICCS reflect a certain normative understanding that expresses a narrow view of the functions of civic and citizenship education (Olson, 2012; Joris & Agirdag, 2019). In this vein, Abs (2013) shows that the ideal type of citizenship portrayed in ICCS is limited to the perception of the "good citizen" as someone who actively engages in political life, while alternative understandings of "good citizenship" – such as liberal and

communitarian ideals of citizenship – are represented to a lesser degree. Another line of normative criticism refers to the involvement of external, international organisations in the development of ICCS and the influence that they are able to exert on the study (e.g., Rutkowski & Engel, 2010).
The second group of criticism contains research that focusses on the theoretical framework that underpins ICCS. Zurstrassen (2011), for instance, criticizes the lack of an overarching theoretical model of proficiency in the study, casting doubt on the utility, applicability and comparability of the ICCS data. Other scholars criticise the theoretical conceptualisation of citizenship ideals (Abs, 2013), the absence of a differentiation between political cognition and social-moral cognition (Biedermann & Oser, 2013), and the insufficient attention paid to the "contestedness" of the concepts and constructs measured in ICCS (Bruun, 2013).
The third category of critiques refers to publications that are critical of the methodology that underlies the design of ICCS and the measurement of its constructs. Some scholars criticise the exclusion of students – or the lack of co-creation methods – in the development of the study (Mason & Delandshere, 2010; Olson, 2012). Other critiques argue that, due to their contested nature, attitudes and dispositions should not be measured by Likert scales or reported as unidimensional scales (Bruun, 2013; Zurstrassen, 2011). Yet other scholars question the appropriateness of ILSA methodology for studying civic and citizenship education (Joris & Agirdag, 2019; Mason & Delandshere, 2010).
The strands of normative, theoretical and methodological criticism of ICCS pose some questions for comparative research into civic and citizenship education in general and ICCS specifically. (1) Are the critiques valid? (2) Are the limitations pointed out by the critiques justifiable given the stated aims of ICCS? (3) Do the critiques point at any actionable improvements that can be implemented in future iterations of ICCS?
The second part of the paper discusses these questions in order to formulate a response to the second research question. When working with specific critiques of ICCS, it is important to keep in mind that ICCS – as all ILSA in education – comes with certain limitations by design. For instance, the necessity of standardisation of study instruments and procedures in order to yield internationally comparable data makes it impossible for ICCS to pay close attention to all idiosyncratic approaches to and unique perspectives on civic and citizenship education that exist in the participating education systems. Likewise, the desire to measure trends in educational outcomes and contexts over time necessitates a certain consistency in the study instruments, which in turn limits the ability of ICCS to incorporate all relevant aspects of civic and citizenship education. In other words, the level of abstraction needed to study and contrast civic and citizenship education at the level of the education system comes with the loss of some degree of detail.
The criticism suggests that the limitations and intentional design choices underlying ICCS are not communicate well enough to various audiences, including the research community. At the same time, the critiques hold some merit with regards to actionable improvements to ICCS that are worth exploring. For example, the theoretical underpinnings of ICCS could be examined more closely in order to establish a more robust link to an adequate model of proficiency and cognition. Similarly, the question whether the educational outcomes measured by ICCS can be considered unidimensional warrants further attention. Moreover, the normative, theoretical and methodological critiques of ICCS shed light on the multitude of research work that can be undertaken in order to complement ICCS results, employing, for example, qualitative methods and co-creation approaches.
By reviewing and categorizing criticism on ICCS (and CIVED) and discussing them in light of the characteristics of ILSAs, this paper aims to identify aspects of the study that warrant further investigation as ICCS evolves. In this way, the paper aims to contribute to the improvement of comparative research into civic and citizenship education and stimulate scholarly debate on this matter.

Keywords: Systematic Literature Review, Civic Education, Citizenship, International Large-scale Assessment

SESSION 13. INITIAL TEACHER EDUCATION FOR SECONDARY SCHOOLS: EVIDENCE AND PERSPECTIVES

ORGANIZER: PEGASO TELEMATIC UNIVERSITY - UNIVERSITY OF FOGGIA
COORDINATOR: ANNA DI PACE – MARILENA DI PADOVA – ANDREA TINTERRI
NOVEMBER 19TH: 2.00 P.M. – 4.15 P.M. {ROOM 4 – RESEARCH 3}

Territorial equity in habilitation pathways: an analysis of the ratio of students/places in grade I and II secondary schools

Marika Lamacchia – Francesco Facciorusso – Maria Concetta Carruba

Introduction

The reform of teacher recruitment initiated by the Prime Ministerial Decree (DPCM) of August 4, 2023, and further detailed in subsequent decrees published in Official Gazette No. 188/2023, has redefined the 60 ECTS qualifying pathways for lower and upper secondary school teachers in Italy. Each region has been allocated a quota of training places, theoretically proportional to local educational needs. In a context marked by chronic teacher shortages (Barbieri & Rossetti, 2022), the rationale behind this distribution has become central to both academic and institutional debates. Previous evidence indicates that imbalances between authorized positions and actual demand undermine both teaching quality and employment stability (ANVUR, 2020; De Santis, 2021). International literature further confirms that classes with fewer than fifteen students yield better learning outcomes (Chingos & Whitehurst, 2011; OECD "Education at a Glance" 2023). Building on this foundation, the present study analyzes the alignment between authorized training positions and regional student populations for the 2023/24 academic year.

Research Scope, Objectives, and Hypotheses

This investigation examines the territorial distribution of the 2023 60 ECTS pathways, with three primary objectives:

- To calculate, for each region/autonomous province and for both school levels, the student-to-training-place ratio;
- To estimate theoretical coverage, assuming a standard of 15 students per teacher (and to test sensitivity with thresholds of 18 and 20);
- To identify significant differences between school levels and among the macro-areas of North, Center, and South.

The hypotheses are as follows: (H1) In certain regions, the student-to-place ratio significantly exceeds the national average; (H2) Upper secondary schools exhibit lower coverage rates than lower secondary schools, as already observed in the 2016–2022 competitive examinations (MIUR, 2023).

Data Sources

The primary sources are official legislative texts: the DPCM of August 4, 2023 (Official Gazette No. 224, 25/09/2023), and subsequent supplementary decrees (Official Gazette No. 188/2023), from which the authorized 60 ECTS training places for each subject and region were extracted. Data on educational demand (number of enrolled students) were obtained from the MIUR Open Data "Student Area" for the 2023/24 school year, aggregated by region and school level. Supply and demand data were integrated into a unified database, indexed by region and school level, from which the relevant quantitative indicators were derived.

Methodology

- *Phase 1 – Indicators*: Calculation of (a) student-to-place ratio; (b) theoretical coverage rate (places ÷ standard requirement).

- *Phase 2 – Descriptive Analysis*: Regional choropleth maps and comparative bar charts illustrate indicator distributions and corresponding 95% confidence intervals.

- *Phase 3 – Inferential Tests*: Normality assessed with Shapiro-Wilk test ($\alpha = 0.05$); if violated, Wilcoxon signed-rank test applied. Comparisons between lower and upper secondary levels conducted via paired t-test (or Wilcoxon), with Welch correction for heteroscedasticity (Levene/Brown-Forsythe test). Differences among macro-areas analyzed using Welch's one-way ANOVA, with Games-Howell post-hoc tests. Analyses conducted in R 4.3.2.

Expected Results and Argumentation

The analysis is expected to reveal that certain southern regions have particularly high student-to-place ratios compared to the national average, indicating a greater shortage of qualifying pathways relative to local demand. Simultaneously, a more pronounced saturation of training places is anticipated in upper secondary schools, consistent with the high demand for teachers in this segment. Overall, the results will provide a quantitative overview of territorial disparities: regions with low theoretical coverage will require corrective interventions. These findings will inform policy decisions aimed at rebalancing training resources and optimizing the allocation of qualifying pathways across the country.

This adaptation maintains the academic tone and structure, referencing relevant policy and research, and aligns with conventions in English-language scholarly writing on education policy and teacher recruitment.

Keywords: qualifying seats; students enrolled; student/seat ratio; territorial distribution

KTUM: a report on the development of a small-scale pan-European standardised test to measure the mathematical content knowledge of future upper secondary mathematics teachers

Ottavio Giulio Rizzo

Introduction

Throughout the world, students' mathematical knowledge is routinely evaluated using written items, both in formative and summative assessment. A similar approach to evaluate teachers' mathematical knowledge is less common. It could be seen as sensitive or even offensive by teachers. A third case, that is somewhat intermediate, is that of pre-service teacher students. At universities, their mathematical knowledge is also commonly evaluated at written exams.

In an international context (with 17 countries), a comparative assessment of teacher students' knowledge was carried out by the TEDS-M study for prospective primary and lower-secondary level teachers of mathematics (Tatto, 2013). A seed project (called "Knowledge for Teaching Upper secondary Mathematics", KTUM) was approved by the 4EU+ alliance to develop a similar approach within four European universities (located in Czech Republic, Denmark, France, and Italy), to develop a shared model of the mathematical content knowledge that is deemed relevant in the participating universities for students close to qualifying for upper secondary — as opposed to primary and lower secondary as in TEDS-M — teaching (Rizzo & Winsløw, 2025).

The four countries involved in the project represent the full gamut of approaches to pre-service teacher training: from five years of Maths Education at Charles University (Prague) to five years of pure Mathematics, followed by a 20 ECTS Practicum, at Copenhagen; from a bachelor in pure Mathematics followed by a master's degree in Maths Education in Paris, to a bachelor in pure Mathematics followed by a master's degree with a concentration in Maths Education and by a 60 ECTS Practicum in Milan. Moreover, the different hiring processes — national competitive contests (Italy and France) Vs. qualifications with schools hiring teachers directly (Denmark and the Czech Republic) — carry an influence on the institutional settings.

The research question

Many studies, both in the US (Hill et al., 2008) and internationally (Krainer et al., 2015), have tried to measure mathematics teacher knowledge and its link to effective teaching (Winsløw, 2024). Early findings show that simply taking more university math courses does not clearly improve student outcomes. Instead, deep understanding of the subject and how to teach it — known as mathematical content knowledge (MCK) and mathematical pedagogical content knowledge (MPCK) — are more relevant (Tatto, 2015). These categories are often assessed through tests, mainly for primary and lower secondary teachers. However, KTUM appears to be the first study examining MCK and MPCK for upper secondary teachers in Europe. This study adapts TEDS-M's MCK subdomains (Table 1) and includes some MPCK items tailored for the upper secondary level.

Given the context, the research question we focus on in this presentation is: what kinds of items are proposed by representatives of European universities, when asked to provide input to an MCK/MPCK test for future upper secondary school teachers? And what are the issues connected with the development of a questionnaire comprising these items?

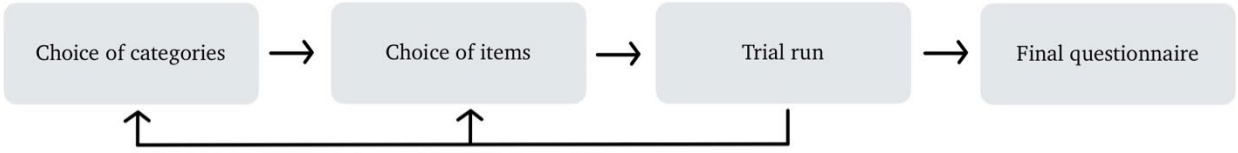
Table 1: MCK framework: subdomains and content areas (adapted from Tatto, 2013, p. 13)

Subdomain	Content Areas
Numbers	Number sets (e.g. rational); Induction; Number Theory
Geometry	Geometric shapes; Geometric measurements; Analytic Geometry
Algebra	Algebraic expressions; Algebraic equations; Vectors, Matrices
Functions	Equations; Algebraic and Transcendental functions
Data and Chance	Data processing and interpretation; Probability
Analysis	Continuity; Limits; Differential calculus; Integral calculus

The design process

The very first step was mapping out the typical path for a student to qualify for teaching mathematics at the upper-secondary level: this is a combination of the national legal framework and curricular choice of the given university. For example, in the case of the University of Milan, we considered a student attaining a bachelor's and a master's degree in mathematics, with latter having a concentration in Maths Education. Notice also that the Italian regulatory framework was in a state of flux during the project, given the reintroduction of a formal teacher training practicum between the master's degree and the national contest. It was expected that the curricula in the four countries differ on points like whether or not synthetic geometry is emphasised or not. It was a surprise to learn that, in the Czech Republic, differential and integral calculus is no longer part of the upper secondary curriculum. Nevertheless, since essentially all the concerned students in the Czech programme will also have taken courses in analysis, and analysis is a major topic in upper secondary school in the remaining three countries, it was agreed that this domain should be part of the questionnaire. As an example of more detailed differences, which affected our selection of items in the end, was whether teacher students could be expected to know the rational root theorem or not. The principle was to include only items that could be solved by knowledge taught in all four programs.

Figure 1: Development cycle of the questionnaire (from Rizzo & Winsløw, 2025)



Each national group initially proposed 5-7 items based on their experience and practice of assessing MCK and MPCK. These items were mostly taken or adapted from exams in teacher-oriented courses, from

national contests and standardised tests (INVALSI tests), etc. The KTUM team then met for three days in Prague, working in mixed groups (people from different universities) to discuss these first items, propose new items to cover missing areas or aspects (especially MPCK), and to fine tune the resulting inventory of items in terms of balance and formulation. Many items were discarded because some national group did not find them appropriate for their students.
The first phase was entirely conducted in English, being the only language mastered by all members of KTUM. The second phase required the items to be translated into the four relevant languages (Italian, Danish, French, Czech): since French is mastered by most members of KTUM, the French translation was used to double check translation issues. One example of these issues is the difference between "dot product" and "scalar product": in English, the latter typically refers to any symmetric bilinear form, while the former refers to the standard scalar product on $\mathbb{R}^n$. In, say, Italian, only the latter expression is used: "dot product" should then be translated as "prodotto scalare standard" (standard scalar product).
The group collaborated on a common LaTeX file hosted on OverLeaf, and special routines to deal with international text were developed in order to minimise the risk of inconsistencies in the different localisations, it was deemed paramount that the mathematical part was not part of the localisation, and that the LaTeX machinery should be transparent enough to be usable by project members with little, or no, previous LaTeX experience (see Figure 2).

```
\en{Find} \it{Trova}\da{Bestem et tal}\cz{Najděte } \fr{Trouver  } $x$
\en{such that }\it{tale che }\da{ så }\cz{ takové, že }\fr{  pour que }
$\left(\sin(1/x)\right)^{-1}$
\en{has 128 digits}\da{har 128 cifre}\cz{má 128 číslic}\fr{s'écrive avec 128 chiffres}\it{abbia 128 cifre}.
```

Figure 2: Example of localised LaTeX code

At the end of the first phase, 14 items (most with several questions) were selected and administered to a group of 17 volunteer students: four from each of Milan, Prague and Paris, and five from Copenhagen. The students did the test in the same room in Milan, and after the test, a thorough follow up discussion with them was organised (also to let them know how to solve all items). The sample of students was far from being statistically significant or random as they were volunteers. All except the Milan students would presumably have been motivated by a paid trip to Italy. Still, the students' performance and assessment of the items was believed to be valuable for the further development of the item inventory.

Initial test items

As mentioned above, 14 items were selected for the first trial: 6 involved the domain Numbers, 7 Algebra, 3 Geometry, 2 Functions, 2 Data and Chance, 3 Analysis. Here, the total is greater than 14, since most items contain several questions which could be in different domains). Considering all questions, 17 involved the domain Numbers, 8 Geometry, 15 Algebra, 8 Functions, 3 Data and Chance, 6 Analysis. For the first trial test, students were given three hours. This allowed to experiment several items, mostly with open response questions. However, for a large-scale test, a shorter test time would be more realistic. Therefore, at any rate, some items or questions had to be dropped.
A main principle was to discard questions that were done correctly by all, or by none (or very few) of the students (so they had little or no discriminatory power). For example, in Figure 3, we have three apparently similar questions which test students' MPCK. The first one was selected by three (out of 17) students, the second was selected by half the students (IT: 1/4; DK: 0/5; CZ: 2/4; FR: 3/4), and the last one by none. We retained to two first options.
The second answer illustrates the mutual dependency of MCK and MPCK: during the follow-up discussion, we learned that the students who had not marked this exercise simply did not know the relationship between logarithms and number of digits (MCK). So MCK clearly influences MPCK. The piece of MCK concerned is useful to circumvent difficult actual calculations and relates to intrinsic and historically crucial aspect of logarithms. The fact that Danish students did not know this might be related to the extensive use of computers in Danish upper secondary school. By contrast, that 3 of 4 French students marked the second exercise could be caused by the fact that French high school students are expected to solve exercises with as little help as possible from technical tools. Thus, differences in (relevance of) MPCK may also lead to differences in MCK.

Emma wants to select exercises that can be solved by using logarithms and their properties. Mark all the exercises that are relevant for her.

- ◯ Find x such that $(\sin(1/x))^{-1}$ has 128 digits.
- ◯ How many digits has 3^{280}?
- ◯ How many digits has the integer part of $\sin(3^{280})$?

Figure 3: A question on choosing exercises (extract)

We expected the questions in Figure 4 to be quite doable for master's level students in Mathematics or Mathematics Education, with the only difficulty being in the writing of a proper proof using one of the various possible arguments. All the Czech students answered the first question of item 13 correctly, while none of the French, one Italian and one Danish did so. However not a single student gave correct answers for the first question of item 14, and no student gave mathematically rigorous answers to the second questions. Both items were discarded.

13. For which natural numbers n does the equation $x^n + x + 1 = 0$ have at least one real solution? Explain your answer.
14. For what natural numbers n does the equation $x^n + x + 1 = 0$ have at least one rational solution? Explain your answer.

Figure 4: Questions that were too difficult

Acknowledgments
Laura Branchetti, Renaud Chorlay, Naďa Vondrová, Carl Winsløw, David Janda, Jakub Michal were active partners in the KTUM project and in the development of the questionnaire. The project was financially supported by the 4EU+ European University Alliance under a SEED4EU+ call.

Keywords: Teacher training, large scale assesment, Mathematical Content Knowledge, international assesments

E-portfolio as a tool for Building Professional Identity in the Internships for Initial Teacher Education

Anna Dipace – Teresa Savoia – Cristiana D'anna – Raffaele Di Fuccio

Introduction
The DPCM of August 4, 2023 has substantially restructured the pathways for initial teacher education and certification, introducing a unified model of 60 or 30 ECTS credits depending on access requirements. At the heart of this reform lies the teaching internship, composed of a direct component (carried out in schools, also called practicum) and an indirect component (laboratory-based and reflective in nature). This integration of theory and practice, the presence of expert supervision, the creation of professional learning communities, and the assessment of training processes through e-portfolios, field observations, and written reflections are essential elements of high-quality initial training (Darling-Hammond, 2017). In particular, the training portfolio is identified as one of the core tools of the pathway: not only as a documentary proof, but as a reflective, evaluative, and guiding device in the professionalization process. This function is specified in Article 7, paragraph 3, and Article 10, letters b and d of the DPCM, and finds confirmation in pedagogical literature that recognizes its transformative value (Carrubba, Giacomantonio, Pilotti, 2022). In line with these principles, a recent study on the e-tivity framework highlighted how digital tools support the documentation of learning and reflective self-assessment (Dipace, Fiore, Savoia, 2023). Unlike purely certifying tools, the e-portfolio represents a dynamic space capable of supporting the construction of the professional self (Grządziel, 2020), reflecting the complexity of professional knowledge and helping the future teacher become aware of their own resources, limitations, and development strategies.

Research Purpose, Objectives, and Hypothesis
This contribution aims to analyze the theoretical and operational framework of the trainee's e-portfolio as adopted by Pegaso Digital University, within the new qualifying pathways established by the DPCM of August 4, 2023. Two main objectives guide the study: on the one hand, to present the digital e-portfolio tool developed by the University to support trainee professional development; on the other, to examine a selection of significant case studies from the 180 portfolios produced in the Umbria region, illustrating how students used the tool to reflect on, document, and design their training experiences. The hypothesis is that, if adequately designed and integrated into the training path, the e-portfolio facilitates the transition from student identity to teacher identity, promoting pedagogical, methodological, relational, and organizational competences in a conscious and reflective manner.
Data Used
The reflection is based on the analysis of the e-portfolio template developed by the University for students enrolled in the 60- and 30-credit pathways, in compliance with current regulations. The portfolio includes: an initial and final skills assessment; thematic reflections on seven dimensions (disciplinary, methodological, inclusive, planning, assessment, relational, and self-evaluation); analyses of critical incidents; lesson plans and field observations; and a final report. A sample of portfolios from trainees of various secondary school teaching subjects who completed the program between November 2024 and April 2025 in Umbria was analyzed anonymously and with consent. Among these, several particularly representative case studies were selected to highlight reflective practices and recurring professional elements in the growth process.
Methodology
The methodology adopted is qualitative with a descriptive-interpretive approach. Content analysis was conducted using an analytical grid based on indicators drawn from the literature on self-directed learning and teacher professionalism. The model used is that of Carrubba, Giacomantonio, and Pilotti (2022), based on three axes: critical reflection, documentation of significant teaching practices, and development-oriented self-evaluation. These axes were also applied to the case studies to assess the emergence of awareness, agency, and professional responsibility.
Results
The analysis aims to yield results regarding the following areas:
- **Reflective Function** – Has the analysis of critical incidents and professional beliefs fostered a reflective attitude? Have the trainees problematized complex situations, identifying alternative methodologies grounded in pedagogical principles?
- **Self-evaluation and Development** – Have the initial and final skills assessments, together with self-compiled rubrics, enhanced awareness of professional development and the identification of areas for improvement?
- **Theory-Practice Mediation** – Have the teaching plans enabled the concrete application of theoretical references, translating them into inclusive operational activities and coherent assessment tools?
- **Professional Identity and Self-direction** – Has the portfolio created a space for identity construction, in which trainees articulated their values, educational style, and future perspectives? Are the results in line with Dipace, Fiore, and Savoia (2023), according to whom the e-portfolio acts as a "formative self-direction environment"?

Conclusions
From an improvement and innovation perspective in teacher education, the discussion on the theoretical and operational framework of the e-portfolio model — together with the analysis of results across four closely interconnected domains—supports a methodological reflection aimed at monitoring and evaluating this key tool for building teaching professionalism. The e-portfolio functions as a true activator of transformative processes (Moon, 2000), underlining the need for replicable models and good practices that can effectively respond to educational needs, fulfill documentary, narrative, evaluative, self-evaluative, certifying, and guiding functions (Sasanelli, 2024), and provide the teacher with "a cognitive compass that directs, regulates, and supports individual learning and career-building strategies" (Pellerey et al., 2018), thereby promoting deep and conscious professional growth.

Keywords: E-portfolio, initial training of teachers, trasnformative education, self-reflective assessment

Can qualification pathways change the methodological choices of in-service teachers? A longitudinal analysis.

Marilena di Padova – Maria Concetta Carruba – Andrea Tinterri – Anna Dipace

Introduction

In recent years, the Italian school system has faced increasingly complex challenges, requiring a profound rethinking of teaching methodologies, greater attention to inclusive strategies, and a critical and effective integration of digital technologies. Teachers' ability to adopt innovative approaches, personalize learning, and create inclusive environments has become crucial to meeting the diverse educational needs of students. In this context, international research highlights how professional development should enhance advanced pedagogical skills, critical and conscious use of technology, and a strong sensitivity to classroom dynamics (Høydal, Finne & Malmberg-Heimonen, 2024). The effectiveness of teacher training and their ability to improve educational practices depend on multiple factors, including demographic, motivational, and professional experience-related aspects (Mpuangnan, 2024). The new training and recruitment system outlined by Legislative Decree no. 59/2017, updated by Decree-Law 36/2022 (Law 79/2022), together with the qualification pathways established by the DPCM of August 4, 2023, aims to strengthen these aspects by providing teachers with tools to integrate active methodologies, educational technologies, and inclusive practices into everyday teaching. The longitudinal evaluation of the impact and effectiveness of these pathways, also in relation to the characteristics of the cohorts involved, is essential to ensure a real impact on professional skills and the overall quality of teaching. In this regard, pathways aimed at experienced teachers—those already qualified in other subject areas or specialized in support activities, enrolled to acquire the additional 30 ECTS qualification (Art. 13 of DPCM August 4, 2023)—assume particular relevance. This sample is of strategic interest as it includes teachers with direct school teaching experience who, during these pathways, have the opportunity to update and rethink their didactic and methodological practices, often deeply rooted in years of teaching. For this reason, understanding the capacity of these pathways to enrich and influence the methodological and planning aspects of these teachers serves as an important litmus test for the effectiveness of the qualification model for in-service teachers.

Object, objectives and research hypotheses

This study examines the results of a survey conducted within the third cycle of qualification pathways offered by Pegaso Telematic University between January and March 2025. The sample consists of in-service teachers with prior experience. The analysis focuses on the preferences expressed regarding the teaching methodologies adopted, inclusive strategies implemented, assessment tools used, and technological resources integrated into daily practice. The general objective is to understand how participation in the training pathway can influence methodological choices and the inclination to innovate teaching among experienced teachers. Specific objectives include: analyzing the evolution of methodological and technological preferences; evaluating openness toward more active, inclusive, formative, and reflective practices; and investigating the redefinition of perceived professional development needs. The research hypotheses are: 1) that participation in the qualification pathways promotes an observable change in the use of teaching methodologies, inclusive strategies, and digital tools; 2) that the pathway stimulates a greater propensity to adopt innovative and personalized assessment practices.

Method or approach

To address the research hypotheses, a pre- and post-intervention questionnaire was developed and administered to participants, including:

- Demographic and professional information (6 items)
- Frequency of use of digital teaching tools (18 items)
- Frequency of use of teaching methodologies (14 items)
- Use of inclusive teaching strategies (5 items)
- Assessment methods used (8 items)
- Perceived competence in instructional design (6 items)

The questionnaire was administered online (CAWI) via Google Forms, anonymously and voluntarily, during the first and last course sessions, on January 23 and March 3, 2025, respectively, through QR code scanning. The questionnaire included a request to create a unique identification code to match pre- and

post-responses from the same participants. The pre-questionnaire collected 1,493 responses, while the post-questionnaire collected 227 responses. Data analysis involves matching pre- and post-responses via the unique code and applying non-parametric statistical techniques for paired samples (Wilcoxon signed-rank test), considering the ordinal nature of the scales used and the need to detect significant differences without assuming normality. Data preparation and analysis were carried out using Google Forms and Jamovi 13.

Results or discussion

The ongoing analysis offers the opportunity to examine the impact of the qualification pathways in promoting more conscious, inclusive, reflective, and technologically integrated methodological choices among in-service teachers. Unlike previous studies, mainly aimed at diagnostic analysis (di Padova, Tinterri, Dipace 2025), the present study seeks to understand how the training pathway can concretely guide teachers' didactic practices and operational preferences. In particular, the investigation focuses on the variety and frequency of use of digital teaching tools, innovative methodologies, inclusive strategies, and formative assessment methods, relating these aspects to demographic and professional variables (age, gender, experience, disciplinary area). Involving already active teachers, the pathway provides an opportunity to investigate whether and how such training experiences can foster a significant shift toward a more active, personalized, and critical approach, contributing to the enhancement of teaching quality and the rethinking of professional development needs. In summary, the present study not only contributes to mapping teachers' methodological and technological preferences but also provides useful insights for designing future qualification pathways that are more targeted and aligned with the challenges of contemporary schooling.

Keywords: Teacher Professional Development, Teaching and learning methods, Longitudinal Analysis, Inclusion

Assessing the Impact of Teaching Qualification Pathways: A Longitudinal Analysis on In-Service Teachers' Self-Efficacy

Andrea Tinterri – Marilena di Padova – Maria Clara Dicataldo – Anna Dipace

Introduction

In recent years, research on the Italian school system has highlighted worrying criticalities, with learning outcomes generally below the European average in secondary education. These signals have led policymakers to revise curricula, teaching methods, and, not least, teacher training. In an increasingly complex and multifaceted school context, international studies emphasize that the development of teachers' competences is primarily a responsibility of school leadership (Skantz-Åberg et al., 2022). To address the growing need for more solid and up-to-date training for Italian teachers, the new system for initial teacher education and recruitment in lower and upper secondary schools, outlined in Legislative Decree no. 59/2017 and amended by DL 36/2022 (converted into Law no. 79/2022), includes a university and academic qualification pathway structured into training, final exam, and final evaluation. The qualification pathways introduced by the DPCM of August 4, 2023, are designed to strengthen pedagogical, disciplinary, and digital competences, supporting targeted and continuous interventions in teaching practice, with differentiated pathways for teachers already holding qualifications in other subjects or with specialisation in support teaching (art. 13). The effectiveness of these pathways is crucial for the success of secondary education reforms; it is therefore essential to obtain evidence to measure their effectiveness and to adapt methodological and didactic choices to the actual needs of teachers.

Object, objectives, and research hypotheses

This study examines the results of a survey conducted within the third cycle of qualification pathways offered by Pegaso Telematic University between January and March 2025. The survey was carried out during the program aimed at teachers already qualified in another teaching subject or specialised in support activities, in order to acquire an additional 30 ECTS qualification (art. 13 of DPCM August 4, 2023). The sample thus consists of in-service teachers with teaching experience. This study focuses on analyzing the extent to which this pathway influences teachers' perceived self-efficacy. Self-efficacy refers to an

individual's belief in their capability to succeed in a specific situation (Bandura, 1997). It is linked to significant educational outcomes, including persistence, enthusiasm, commitment, and instructional behavior, and it has a predictive role in the teacher's ability to generate learning outcomes (De Biasi, Dominici, Patrizi, Capobianco, 2014). The research hypotheses are: (1) participation in the qualification pathways improves experienced teachers' perceived self-efficacy; (2) participation affects teachers' initial perception of their own training needs.

Method or approach

To address the research hypotheses, a pre- and post-intervention questionnaire was developed and administered to the participants, including:

- Demographic and professional information (6 items)
- Perceived levels of student performance, motivation, behavior, and cooperation (Cantoia, Clegg, Tinterri, 2023)
- The SAED scale, Italian version of the Teacher Self-Efficacy Scale, consisting of 24 items rated on an 8-point scale (De Biasi, Dominici, Patrizi, Capobianco, 2014)
- Perceived training needs (6 items)
- Expectations regarding training (6 items)

The questionnaire was administered online (CAWI) via Google Forms, anonymously and voluntarily, during the first and last sessions of the course (January 23 and March 3, 2025), through a QR code scan. Participants were asked to create a unique identification code, allowing the pairing of pre- and post-responses. The pre-questionnaire collected 1,493 responses, while the post-questionnaire received 227 responses. Data analysis involves matching pre- and post-responses using the unique code and applying non-parametric statistical techniques for paired samples (Wilcoxon signed-rank test), considering the ordinal nature of the scales and the need to detect significant differences without assuming normality. Data preparation and analysis were conducted using Google Forms and Jamovi 13.

Results or discussion

The ongoing analysis offers an opportunity to explore self-efficacy perceptions among teachers in training and to develop interpretative insights into the effectiveness of the didactic model. Unlike previous studies that were purely diagnostic (di Padova, Tinterri, Dipace, 2025), this study aims to examine the impact of the training pathway on teachers' self-perception. Specifically targeting experienced teachers who have the opportunity to enrich their competences through this program, the study will help determine whether these pathways, as currently structured, can positively impact teachers' perception of their instructional efficacy and perceived training needs.

Keywords: Teacher Professional Development, Teacher Self-Efficacy, Longitudinal Analysis, Teaching Qualification Pathways

Multidisciplinary Centres and Professional Agency: Towards a New Reflective Education

Viviana Vinci – Pierangelo Berardi – Carmela Paladino

Introduction

Preparing the next generation of teachers—capable of navigating the complexity of disciplines, active learning methodologies, and digital technologies—has become a strategic priority for Italian higher education. Teacher education today cannot be reduced to the mere transmission of knowledge; rather, it must foster a cultural transformation that enhances individual agency, professional awareness, and reflective learning. In this regard, international *Teaching and Learning Centres* (TLCs) have proven particularly effective when configured as agents of cultural change, promoting collective processes of pedagogical and professional innovation (Perla, Vinci & Agrati, 2025).

In line with this vision, the recent reform of initial teacher education—introduced through the revision of Article 2-bis of Legislative Decree 59/2017, as amended by Article 44 of Decree-Law 36/2022, and formalized by the Prime Minister's Decree of August 4, 2023—assigns a central role to universities in the

design of the new 60-credit (CFU/CFA) teacher qualification programs. These programs establish *multidisciplinary centres* as strategic nodes for integrating subject-specific pedagogy, professional workshops, and school-based internships, within a training paradigm grounded in evidence.

A comparative glance at international contexts reveals well-established operational models. For over thirty years, TLCs in North America, the United Kingdom, and Australia have combined pedagogical research, faculty development, and curricular innovation, contributing to the spread of a teaching culture grounded in impact data (Haras et al., 2017; Beach et al., 2016; Czerniewicz, 2021). The literature documents their effectiveness through principles of good practice (Sorcinelli, 2002) and measurable outcomes in both student learning and faculty development (Wright et al., 2018).

Despite contextual differences, the emerging Italian *multidisciplinary centres* share a similar mandate with TLCs, aimed at enhancing teaching quality, supporting professional growth, and implementing evidence-based evaluation. The interaction between existing TLCs within universities and the newly established centres can foster significant synergies—particularly in the management of blended modules, reflective practice labs, and systematic data collection on learning. Established frameworks such as the *CTL Matrix* of the ACE-POD Network (2018), the French *IUFM ateliers*, and the *DidaSco* prototypes (Perla, Vinci & Agrati, 2017), offer transferable metrics and tools. Notably, the DidaSco model supports the design and classroom implementation of didactic "prototypes" that integrate research, recursive reflection, and authentic assessment: a coherent proposal aligned with the mission of multidisciplinary centres and readily applicable to reinforce the university–school alliance through evidence-based practices.

Although the new decree assigns multidisciplinary centres a mandate analogous to that of TLCs—focused on teaching, internships, and evidence-based evaluation—their operational configuration remains largely undefined. Article 5 of the Prime Minister's Decree assigns them broad responsibilities, from managing the qualification pathways to coordinating tutors and technological infrastructure. However, it does not specify uniform criteria regarding staffing, digital resources, or reward mechanisms based on employment outcomes. This open model risks generating disparities across institutions, with potential implications for equity in training provision, as highlighted by TLC literature. The study presented here seeks to provide empirical evidence in support of consolidating these multidisciplinary centres, contributing to the identification of shared practices and replicable quality standards. Situated within the broader debate on the governance of teacher education as a collaborative ecosystem among universities, lead schools, and mentor teachers, this research values the training experience, didactic practices, and ideas for transferability. Within this framework, the DidaSco model—based on the co-design of classroom-tested teaching prototypes accompanied by structured feedback—emerges as a practical reference tool to reinforce quality and coherence across teacher qualification pathways.

Research Aim, Objectives, and Hypotheses

The analysis focuses on sixty individual narrative self-evaluations, voluntarily submitted by participants enrolled in the first cycle of the 60-CFU teacher qualification programme at the University of Foggia. These reflections were produced within the transversal course *"Design and Evaluation of Teaching and Learning Processes"*, delivered entirely online. The sample represents a significant subset of the total cohort of 412 participants. The texts were written according to a shared protocol, which invited the trainees to describe their learning experience, identify key strengths and critical issues, and formulate proposals for transferability to their future teaching practice.

The study aims to achieve three interconnected objectives. First, it seeks to reconstruct the emotional dimension that accompanies the training experience, with particular attention to the emergence of motivational drivers and logistical obstacles. Second, it intends to identify teaching practices perceived as effective or in need of improvement, in order to highlight good practices that may be replicable within the framework of Teaching and Learning Centres. Finally, it aims to map the ideas of transferability developed by the participants, assessing their alignment with co-design and authentic assessment models promoted in the literature on faculty professional development.

Data Used

The dataset consists of sixty written narratives, amounting to a total of 62,583 words, with an average of 1,043 words per text (SD = 148). Each self-evaluation was voluntarily submitted in extended form and developed in accordance with a structured protocol encouraging reflective writing focused on emotional experiences, perceived strengths and weaknesses of the training, and potential applications to future teaching practice. The structural consistency of the texts, combined with their lexical richness and thematic variety, makes the corpus particularly suitable for qualitative analysis aimed at identifying significant thematic patterns.

The investigation yielded a layered and multifaceted representation of the training experience, highlighting both recurring perceptions and emerging systemic issues, as well as evolving design orientations. As such, the corpus provides not only a deep insight into the subjective dimension of professional learning but also a valuable empirical resource to inform the strategic development of multidisciplinary centres, which are tasked with ensuring the quality, coherence, and sustainability of new qualification pathways.

Methodology

A qualitative approach was adopted, grounded in an interpretive framework structured around four analytical categories: strengths, weaknesses, future perspectives, and perceived risks. This schema was adapted—heuristically reformulated—from the SWOT analysis model (Dyson, 2004), and applied to the sixty final narratives submitted voluntarily by course participants at the end of the training. The aim was to extract, through the analysis of these reflective accounts, key elements capable of informing the future design of teacher qualification pathways, in a perspective of continuous improvement and pedagogical coherence.

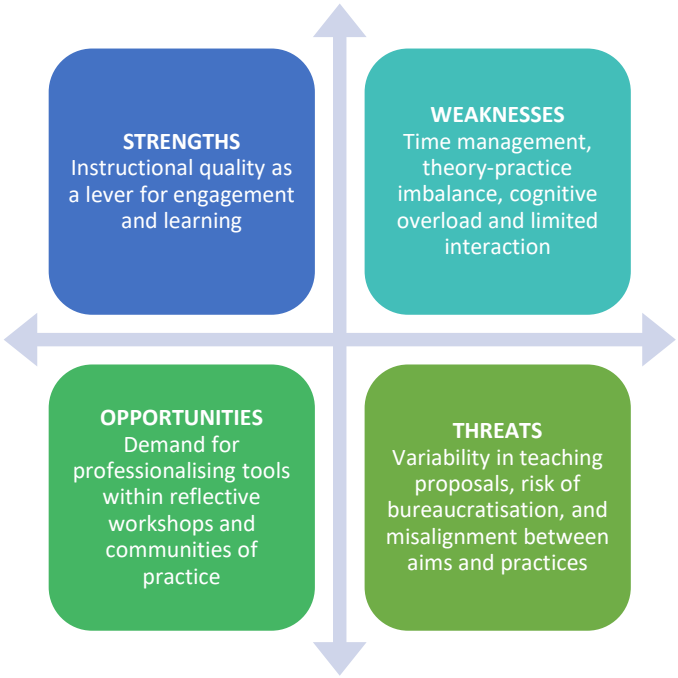


Fig. 1. Summary of the main themes emerging from the qualitative analysis of participants' self-evaluations.

The corpus, consisting of a total of 62,583 words, was first converted into a textual format and subsequently organized into a spreadsheet to facilitate an initial exploratory review. The analysis unfolded through a sequence of iterative phases: a preliminary immersive reading aimed at gaining a holistic understanding of the training experience; a second phase dedicated to the semantic segmentation of relevant content; and finally, an interpretative synthesis phase, which allowed the recurrent themes to be mapped onto the four previously defined analytical categories.

Special attention was given to the intersection between individual expressions and recurring patterns, with the aim of identifying not only explicitly formulated evaluations but also latent resonances pointing to shared educational needs. The adopted method sought to highlight the reflective and transformative dimension of writing, treating the narratives not merely as a source of empirical data, but as spaces for the emergence of metacognitive awareness.

The use of the fourfold framework enabled the representation of the diversity of lived experiences within an interpretive structure that is both flexible and conceptually rich. This heuristic tool proved to be functional both as a diagnostic lens and a design-oriented strategy. In this regard, it supported the refinement of training standards, the enhancement of instructional quality, and the operational sustainability of future multidisciplinary centres.

Results

The qualitative analysis of the participants' narratives revealed four main interpretative dimensions that systematically reflect trainee perceptions and offer valuable guidance for the design of teacher qualification pathways.

The first dimension relates to the quality of instructional delivery, which was unanimously recognised as a key factor in fostering engagement and learning. Clarity of explanation, empathy, and timely feedback were

particularly valued and associated with a positive educational experience. The second dimension highlights organisational challenges, notably the imbalance between theory and practice, which impacts cognitive load and limits the activation of interactive dynamics.
The third dimension points to a widespread demand for professionalising tools, including micro-teaching sessions, simulations, and shared rubrics. These were explicitly requested in workshop formats to support reflective practice.
The fourth dimension concerns a broad convergence around transferability, with active methodologies, feedback technologies, and formative assessment emerging as key components. These align with international best practices and the mandates assigned to multidisciplinary centres. Strengths focused on appreciation for the instructional approach, characterised by communicative efficacy and relational sensitivity, seen as a driver of motivation. The use of innovative and transferable pedagogical tools, such as assessment rubrics and authentic tasks, was also highly valued.
Conversely, weaknesses were mainly linked to content management: excessive theoretical density and limited integration with practical activities reduced teaching effectiveness and peer interaction. With regard to opportunities, reflections converged on the demand for more engaging, context-based pedagogical tools. Participants called for the systematic introduction of micro-workshops, teaching simulations, structured feedback, and peer observation practices within a cooperative and professionalising learning environment. The intersection of design, reflection, and assessment emerged as a strategic axis for shaping initial teacher education capable of influencing real classroom practices.
Identified threats were predominantly systemic in nature, raising critical questions about the governance and sustainability of training programmes. The lack of common standards for structuring multidisciplinary centres, high variability among universities, and the risk of bureaucratisation of assessment practices were seen as destabilising factors that could compromise the alignment between stated goals and actual outcomes. A recurring concern in many narratives was the fear that training quality might become subordinated to a logic of compliance, driven by standardisation rather than the actual needs of schools.
The evidence collected outlines a coherent and multifaceted view of training needs, expressed through clear strategic priorities: the centrality of teaching quality, a rebalancing between theory and practice, the promotion of reflective tools, and the valorisation of transferable practices. This interpretive map, built through thematic categorisation along analytical polarities, functions as a heuristic device for delineating operational pathways and supporting the evolution of multidisciplinary centres toward an integrated, reflective, and evidence-informed model.
The analysis offers not only a detailed portrait of the trainees' learning experiences but also a strategic horizon capable of guiding the planning of future multidisciplinary centres. The findings confirm the coherence between subjective perceptions, systemic challenges, and established theoretical references, outlining a demand for authentic experiences, reflective tools, and effective assessment practices. In this view, innovation resides not merely in content delivery but in the training system's capacity to generate professional awareness and collective transformation.
Established models such as *DidaSco* and international *Teaching and Learning Centres* offer consolidated operational frameworks that can inform the design of qualification pathways based on active methodologies and authentic assessment. Additionally, there is a strong need to foster meaningful dialogue across the disciplines involved in initial teacher education to build shared languages and practices. In this direction, multidisciplinary centres may become catalysts for systemic innovation, drawing inspiration from the constructs of *Signature Pedagogies* and the *Scholarship of Teaching and Learning (SoTL)*, which conceptualise university teaching as a reflective, rigorous, and transformative research domain.
Only an ecological and transdisciplinary vision of teacher education can anchor educational quality in real contexts, translating pedagogical principles into operational tools capable of producing both impact and sustainability (Perla & Vinci, 2022). Multidisciplinary centres, entrusted with ensuring quality and coherence, must take on a proactive role in educational innovation by translating empirical analysis into shared operational standards. Through the systematic adoption of consolidated practices, it will be possible to respond equitably and sustainably to the challenges posed by the Prime Ministerial Decree of 4 August 2023, avoiding inter-university variability from undermining the transformative potential of the new qualification pathways. If properly designed, multidisciplinary centres can evolve from merely organisational structures into strategic nodes of systemic innovation, capable of ensuring quality, equity, and coherence across the entire training system.

Keywords*:* Initial teacher education, Multidisciplinary centres, Reflective learning, Evidence-based design

SESSION 1. TERRITORIAL DISPARITIES IN EDUCATION: A LONG-STANDING ISSUE ADDRESSABLE THROUGH NEW DATA AND TOOLS

ORGANIZER: UNIVERSITY OF MILAN BICOCCA – CATHOLIC UNIVERSITY OF MILAN
COORDINATOR: GIANLUCA ARGENTIN – ELISA MANZELLA
19TH NOVEMBER: 4.30 P.M. - 6.30 P.M. {AULA 5 – RESEARCH 4}

New evidence on math summer learning loss from Italy

Giovanni Abbiati - Gianluca Argentin - Giulia Assirelli - Rosaria Lumino

Introduction

Summer learning loss may significantly contribute to both the low efficacy of school systems and the perpetuation of social inequalities throughout students' academic careers. While extensive research on the intensity and impact of summer learning loss has been conducted in the USA, studies in Europe have been more limited, primarily focused on Central and Northern Europe (Paechter et al., 2015; Meyer et al., 2017; Shinwell & Defeyter, 2017; Holtmann & Bernardi, 2019).

Existing research underscores the potential importance of summer learning loss, though findings remain somewhat inconsistent. This inconsistency likely stems from several factors: differing definitions of summer learning loss, varied estimation methods, attention to distinct school grades or subjects, and comparisons across national school systems with differing summer break lengths, holiday management practices, and access to summer opportunities (Dujardin et al., 2022; Gierczyk & Hornby, 2023; Stewart et al., 2018).

Summer learning loss has been scarcely studied in Italy; only a few studies have been conducted, all based on small and localized samples (Sabella, 2014; Azzolini et al. 2023). This lack of national evidence is surprising, as Italy offers a uniquely relevant context for investigating the effects of summer break. Italy has one of the longest summer breaks globally, coupled with a strong tradition of assigning substantial homework to be completed during this period, and a long-standing habit of calling families to support their children for this task. Furthermore, a national assessment is conducted every year before summer, covering the entire student population at the end of second, fifth, eighth, and twelfth grades.

Research Questions

Our contribution aims at providing new evidence on the summer learning loss in Italy, in order to understand: i.to what extent does summer learning loss intensify under institutional conditions like those in Italy; ii. whether these conditions disproportionately impact students' learning based on their cultural and migratory backgrounds and according to the local contexts where they spend their summer break.

We focus specifically on primary school, particularly in mathematical learning, as early losses in this subject are harder for families to mitigate and may have significant long-term consequences. Italian students already display notable gaps in mathematical skills in international comparisons, and a low percentage pursue STEM fields. By investigating these dynamics, our study aims to shed light on the potential influence of extended breaks on math skills and the broader educational implications for Italy.

Data and Methods

We collected data on math achievement for third graders in a self-selected sample of 73 schools across two Italian provinces, Milan and Naples, which have markedly different socio-economic contexts. At the start of the current school year (2024/2025), we administered a math assessment along with a questionnaire to all

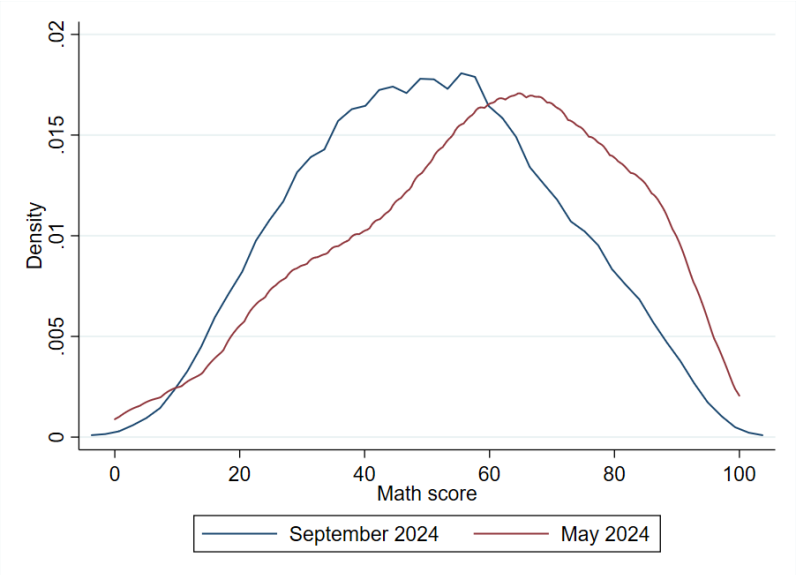
classes (180) within the selected schools. This study leverages the fact that the same students took a similar math test as part of the National School Assessment (INVALSI) in May 2024, at the end of the 2023/2024 school year, while they were in second grade. We successfully matched the May and September scores for 2,515 students out of an initial sample of 3,308. Matching issues arose primarily due to student absences during either the May or September assessments or a designation of needing special support, as scores for these students are not included in the dataset. In this analysis we estimated summer learning loss in mathematics by comparing normalized scores on a 0-100 scale, allowing the summer learning loss to be interpreted directly in percentage points.

In addition, the research team conducted a questionnaire with the math teachers of the selected classes (90% response rate) and with the students' parents (50% response rate). The teachers' survey explored topics such as teaching practices and the amount of homework assigned, both during the school year and over the summer. The questionnaire for parents was designed to gather more granular information about the family's socio-demographic background, as well as parenting styles throughout the school year and summer break. These data have been then successfully linked with students' scores.

Main findings

On average, students scored 7.8 points lower in September than in May, a figure that remains consistent even after excluding outliers (the 1st-5th and 95th-100th percentiles of the loss distribution). Approximately 64% of students experienced a decline in learning, 4% maintained stable scores (±1 point), and the remaining 32% showed improvement. This learning loss, with an effect size of -0.35, is also evident when comparing the distribution of math skills before and after the summer break (Figure 1).

Fig.1 – Math skills before and after the summer break



Overall, summer learning loss appears relatively consistent across variables such as sex, migratory background, birth month, and the number of books at home (used as a proxy for cultural resources). The most important variable correlated with the summer learning loss is previous achievement, that is, the level of competences exhibited before the summer. Quite surprisingly, competences in may correlate strongly, and positively, with the learning loss, indicating that that the most proficient students lost the most. Moreover, we observed a 2.7-point gap between students in Naples and Milan, favoring the latter. We inspected whether this difference was due to a higher prevalence of disadvantaged families in the province of Naples or to a higher proportion of students enrolling one year advance to primary school (a phenomenon pretty common in Southern Italy). Our findings indicate that this North-South difference is not driven by these sociodemographic factors, suggesting that summer learning loss may vary based on local opportunities available to children during the summer.

Further analysis will explore data on parental involvement during the summer and the volume of summer homework assignments to identify the factors driving this regional gap. By the time of the conference, these preliminary findings will be expanded with i. additional data on school environments, ii. math teachers, and—for approximately half the sample—iii. parental information. In addition, we will able to provide a

more precise measure of learning loss based on a Rasch correction of the scores collected in September, and on their anchoring to the May test, currently under development.

Keywords: Summer learning loss, summer slide, math achievement, inequalities, primary education

Education in Tuscan Inner Areas

Silvia Duranti

Introduction

Being a student living in an inner area often means attending small schools, sometimes multigrade classes, with teachers working on temporary contracts or ready to transfer schools at the earliest opportunity. Later in the school career, it may also mean having fewer options when choosing a study path or having to travel far to attend one aligned with personal interests. It also often means achieving below-average learning outcomes and being more likely to finish secondary school without the skills needed to enter the labor market or, more broadly, to face adult life.

Objective

The aim of this work is to analyze the peculiarities of the inner areas of Tuscany in the field of education, considering: 1) the availability and characteristics of the school offer, 2) student skill levels, and 3) the propensity to pursue tertiary education.

Data

The analysis draws on different data sources:

- The first part uses data provided by the Tuscany Region on the number of students, classes, and schools per municipality.
- The second part is based on INVALSI individual-level data from the 2022–2023 school year, covering all levels of education. Thanks to collaboration with INVALSI, a variable indicating whether a student belongs to an inner area of Tuscany was added to the dataset. To ensure school (and thus student) anonymity, INVALSI only provides territorial identifiers for areas containing at least three schools per level. For inner areas, three easily identifiable schools were removed from the dataset to protect anonymity.
- The third part uses ISTAT data on university enrollment rates, by municipality of residence.

Methodology

The classification of Tuscan municipalities as "inner areas" is based on the national mapping approved for the 2021–2027 EU programming cycle. This mapping considers the travel time to a group of three essential services: upper secondary schools, hospitals with emergency services (DEA), and railway transport of at least "silver" level (according to RFI classification). Municipalities that have all three services are classified as "hubs" (or "inter-municipal hubs"), while others are classified based on travel time to the nearest hub as "belts" (within 27 minutes) or inner areas in three tiers: intermediate (27–40 minutes), peripheral (40–66 minutes), and ultra-peripheral (over 66 minutes). Out of Tuscany's 273 municipalities: 13 are hubs, 4 are inter-municipal hubs, 92 are belts, 164 (60%) are classified as inner areas: 67 intermediate, 80 peripheral, and 17 ultra-peripheral. These inner areas cover 67% of Tuscany's territory (over 15,000 km^2) and represent about 24% of the population — nearly 880,000 residents.

In the context of EU structural funding, inner-area municipalities are split into two groups:

- The first includes those that are part of the so-called "Integrated Territorial Strategy of the ERDF" (commonly referred to as SNAI areas), which receive specific funding. To be included, municipalities must be classified as peripheral and/or ultra-peripheral, show consistently negative demographic trends, have inter-municipal coordination (e.g., Unions of Municipalities), and not be too large or populous. For 2021–2027, six such areas were identified: three in the northern Apennines and three in central-southern Tuscany. These areas benefit from national resources (via the budget law), dedicated ERDF funds (OP 5.2), additional funds from other EU and national programs (FSE+, FEASR, FEAMP, FSC), and local sources (including the NRRP).

- The second group includes municipalities classified as inner areas (usually intermediate) but not part of SNAI. These areas still benefit from incentives and are guaranteed at least 30% of the region's total resources, coming from multiple funding streams.

Discussion

The analysis of the school provision in Tuscany's inner areas reveals a generally good distribution across all educational levels and territories. Although the spread of schools among municipalities varies depending on school level and remoteness, all SNAI-defined inner areas in Tuscany have at least one school for primary, lower secondary, and upper secondary education. This is mainly due to the remoteness and poor connectivity of many of these areas, which makes it necessary to maintain access to education — especially in the early years — even when student numbers fall below the legal thresholds.

Maintaining a widespread and accessible school network is often made possible through multigrade class structures, particularly common in SNAI areas for primary and lower secondary schools. Despite this effort, SNAI inner areas still have smaller average class sizes than both other inner areas and non-peripheral ones. Moreover, even where schools are present, the full-time offering is more limited: only 52.5% of primary students in inner areas are enrolled in full-day programs, compared to 57% in non-inner areas.

Upper secondary education is also present in all inner areas of the region, but the total number of enrolled students is lower than for the earlier stages. While primary and lower secondary schools in SNAI areas enroll about 12% of Tuscan students, upper secondary schools in these same areas attract only 8% of the regional total. This is likely due to the more limited variety of programs, pushing students to commute to more central areas to pursue their personal interests or talents.

From the perspective of acquired skills, INVALSI test results show that students in inner areas tend to maintain regional-average performance levels through the end of primary school (grade 5). However, achievement gaps emerge in lower secondary (grade 8) and widen further by the end of upper secondary education, where their scores are significantly below those of peers in less remote areas. This trend is not surprising, as learning is a cumulative process, and gaps linked to individual, school, or territorial factors tend to increase over the course of a student's educational career. Supporting this interpretation, Conti et al. (2015) found that in the early years of schooling, attending schools in small and less urbanized municipalities may be an advantage. This is due to closer ties between school, families, and local institutions, allowing for more responsive scheduling, extracurricular offerings, and support services, fostering a positive learning environment. In later stages, however, achievement gaps may arise due to less stimulating school environments, fewer resources, and the fact that more capable students are incentivized to transfer to better-equipped schools in central areas.

Low learning levels throughout a student's academic path often lead to explicit dropouts. In other cases, students graduate from high school but lack the skills necessary for autonomous, informed participation in society or for successful integration into the labor market — a phenomenon known as implicit dropout. This "grey area" is often underreported in official statistics and may fall outside the scope of targeted support policies.

While explicit dropout can only be calculated at the regional level, implicit dropout rates can be assessed using INVALSI data at the level of inner areas. In these areas, the percentage of students finishing upper secondary education with skills below INVALSI proficiency thresholds is 8.9%, compared to 6% in other areas and 6.5% region-wide. However, it's important to note that this data refers to students attending schools in inner areas — not necessarily residents and may reflect a self-selection effect: high-performing students in peripheral areas often choose to attend schools in more central locations.

While students living in inner areas may be able to complete upper secondary education locally - albeit with fewer study paths available - those wishing to pursue tertiary education almost always need to move to larger urban centers. This need clearly influences university enrollment rates: while 40% of young people in non-peripheral areas enroll in university, this share drops to 35% in non-SNAI inner areas and to 33% in SNAI areas.

Keywords: Inner areas, Invalsi data, skills levels, education choices

Measuring Grade Overage by Exploring Its Applicative Potential and Interpretative Dimensions: A Methodological Proposal

Sibilla Maria Filippi – Lorenzo Luatti

Introduction

The expression "dispersione scolastica" refers to the concept of "school disengagement" and so, to the risk of early school leaving and drop outs. It encompasses a range of issues related to students' failure to complete their education successfully or adequately. (Colombo 2010; Biondi Dal Monte & Frega 2024). An initial conceptual framing defines it as the set of processes in an education system, which sees students to be above age for their grade/class (mainly due to grade/class retention) or dropping out from school entirely. Since the 2000's thee Ministry of Education has referred to this as "explicit early school leaving" (encompassing the broader meaning) namely" a set of phenomena that cause a student's formal education to slow down; violations of compulsory education requirements; early school leaving during or at the end of the school year at various levels of compulsory and post-compulsory education before obtaining a diploma" (MPI 2000). A crucial indicator, upon closer examination, is students' being above age for their grade/class, not only because it includes data on grade/class retention, but also because it increases the risk of dropping out, especially when it occurs in an early age (i.e., during the first cycle of education – 6-13 years old). Moreover, it is correlated with "implicit" school disengagement/early school leaving, since - as INVALSI notes — "students who have fallen behind at least one year achieve, at the end of lower secondary school, significantly lower results than the average student in all basic competencies" (INVALSI 2022). At the end of upper secondary education, they achieve scores on average 10.9 points lower — equivalent to over one year of presumed learning (INVALSI 2024).

While there is extensive literature on the factors that increase the risk of students being above age for their grade/class (Rumberger & Lim 2008; Fondazione Agnelli 2011), there is a notable lack of studies that quantify it, even using general data —is rather sparse and limited to research produced exclusively by institutional sources (MPI 2008; MIM 2024), which the literature on early school leaving often references.

Subject, Objectives, and Research Hypotheses

It is therefore important to consult the available open data and formulate a proposal that, grounded in rigorous methodological criteria (Coe, Waring, Hedges & Ashley 2021), enables researchers to calculate the rate of grade overage (mainly due to grade retention), exploring its practical applications and interpretative dimensions—from the micro level (each school year and each school, from grade 1 to grade 13) to the macro level (municipal, provincial, regional school systems, national geographic areas, Italy), in a diachronic-longitudinal perspective.

It will also be necessary to quantify other measures connected or correlated with grade overage (such as the incidence of grade retention, transfers between upper secondary schools, and the presence of students with non-Italian citizenship). Finally, the implications for educational policies and teaching practices must be illustrated.

But first of all, how should grade overage/school delay be interpreted? High rates of delay may indicate stricter selection criteria within a school, poorer student preparation, or rather a stronger tendency to remain in the school system until the completion of upper secondary education. Similarly, low rates are not only a reflection of adequate student engagement, but may also indicate a greater ease of dropping out.

The definition of grade overage/school delay (and its precise quantification: 1, 2, 3 or more years) can provide further insights into the "ambivalent" relationship between explicit and implicit school disengagement (as often observed by INVALSI; see also Rumberger, 2004; Dustmann & Van Soest, 2008), the positioning of schools (particularly upper secondary schools) within the local education system, and the

interconnections and interdependencies that are produced and reinforced within the school system as a result of disengagement dynamics present in various institutions (and enacted by them, consciously or not). We will address these questions and research hypotheses using a methodology for extracting, processing, and analyzing school-related data, aimed at observing disengagement phenomena at the local level. This methodology was developed and tested by Oxfam Italia in five school systems (Genoa, Turin, Savona, Vercelli, and the province of Arezzo) (Luatti/Oxfam 2024). This contribution offers a synthesis, with practical examples, of what the authors have presented in a forthcoming scientific publication (Luatti & Filippi 2025).

Data Used

The data sources used in this contribution are as follows:

A. *"Portale Unico dei Dati della Scuola" (PUDS)*: provides national datasets in historical series (since the 2015/16 school year), allowing for diachronic-longitudinal analyses of public and private schools (government-accreditated), from preschool to both educational cycles (6-18 years old). The data are disaggregated by school year, single school, and territorial level.
B. *"Scuola in Chiaro" (SC)*: does not provide open format data. For this contribution, we will present selected case studies focused on specific Italian cities and regions.
C. *"INVALSI Statistical Service" (SSI)*: for the purposes of this contribution, we will use data from the following datasets (2024 and historical series): "General Report," "Provincial Level Report," and "Combined Report on Dropout and Excellence."

Specifically, the data used include:

- Demographic and enrollment data of the student population: number of students, number of classes, number of schools, students by gender, and students with non-Italian citizenship (PUDS);
- Data on school pathway discontinuity and school implicit disengagement: particularly grade overage (PUDS); levels of basic competence (adequate/inadequate) across different grades (SSI);
- Data on school concentration: schools with a high concentration of students with non-Italian citizenship (PUDS).

Method or Approach

The analytical method is quantitative, based on the extraction, processing, and interpretation of school data obtained from the above-mentioned portals. This contribution will guide researchers through the various steps required to calculate grade overage and relate it to other data in order to provide a comprehensive understanding of the territory under investigation. The proposed interpretative framework aims to highlight:

1. The trend of grade overage across the 13 school grades within the national system, from the first year of primary school to the fifth year of upper secondary school, with a focus on the critical transition between the two cycles (6-13 and 14-18). This will be analyzed in relation to data on overcrowding in first-year classes, student transfers in and out of upper secondary schools, compliance with guidance recommendations, and more.
2. The uneven distribution of grade overage across different upper secondary school tracks, influenced by both entry-level and in-progress factors. For example, early grade overage (developed during lower secondary or even primary school) is shown to influence and shape students' educational choices. Bidirectional flows of student transfers during or at the end of the school year often result in the redistribution of overage students between tracks and schools.
3. The relationship between the incidence of grade overage and students with non-Italian citizenship (NIC), particularly in schools with a high concentration of NIC students.
4. Possible applications and interpretations using implicit school disengagement data (INVALSI).

Results. Implications for Educational Policies, Practices, and Systems

The school data analysis pathway proposed in this contribution, illustrated through selected examples, aims to raise awareness and deepen understanding within local contexts regarding what schools "do" and "produce." It enables the emergence—based on data—of certain disengagement dynamics embedded within the school system and helps stimulate dialogue between schools, the broader educational system, and local communities, contributing to the redefinition of priorities and actions. In relation to several dimensions that can be quantified through the mentioned portals, awareness and public debate remain insufficient. Consider, for example, the quantitative and qualitative aspects of grade overage and its potential impact on students' educational trajectories, the phenomenon of student transfers in upper secondary education, and the need to build stronger bridges (currently fragile or absent under existing regulations) between lower and upper secondary schools.

School data-based analysis also allows for the quantification—and critical reflection—of significant and persistent phenomena of educational channeling, which are partially addressed in this contribution. These phenomena manifest in upper secondary schools whose student composition is strongly shaped by the regularity or irregularity of educational pathways, both at entry and during progression.
The proposed insights highlight the need for interventions at multiple levels:

- Individual student level, through personalized and adequate support, especially for those most at risk of disengagement;
- Classroom level;
- School or school network level;
- Local and national education system level.

The heuristic potential of school data-based analysis—and the benefits it can bring—could be significantly enhanced by the development of software and applications that make data extraction and processing more accessible and user-friendly. Additionally, the availability of open format data for a range of variables that are currently not easily accessible through the two mentioned ministerial portals would further support this work.

Keywords: School disengagement, early school leaving, grade overage, students with migratory background

Unequal Educational Geographies: Mobility within the Milanese School System

Marta Cordini – Andrea Parma – Costanzo Ranci

School segregation refers to the concentration of a group defined by certain characteristics—such as socioeconomic status or ethnic background—in specific schools. It reflects educational inequalities, but it is more than that. It is both a consequence and a cause of the dynamics of reproduction of existing social structures, as well as of class and territorial advantages and disadvantages. The increasing presence of foreign students in Italian schools (Cordini & Parma 2025) and the growing child poverty (ISTAT, 2024) have brought this issue to the forefront in our country, albeit with a significant delay compared to other contexts (such as the United States, the United Kingdom, or France).
While school segregation can be observed in its distribution and intensity within schools, its causes and consequences extend beyond them. They are particularly linked to both the characteristics of the family of origin—as supported by social reproduction theories rooted in Bourdieu's theoretical approaches—and to the characteristics of the territory in which individuals and schools are located, as well as the overall functioning of the school system. Family agency, territorial features, and the institutional framework significantly influence the intensity, form, and geography of school segregation in each urban context. Consequently, family and territorial resources play a central role in shaping the educational opportunities available and accessible to students, creating substantial gaps already upon entry to primary school. In terms of effects, the concentration of disadvantaged groups within individual classes or schools fosters mechanisms that reproduce social disadvantage, such as downward assimilation (Portes & Rumbaut, 2001), further educational impoverishment through peer effects, and—last but not least—raises issues concerning learning outcomes and performance.
Recent studies and meta-analyses have highlighted a relationship between school characteristics and learning opportunities, showing that students tend to benefit academically from attending schools with more advantaged peers, both economically and academically (van Ewijk & Sleegers, 2010). In short, students' individual backgrounds influence their performance (OECD 2019; 2023), while school characteristics—including socioeconomic and ethnic composition—can affect both achievement levels and the relationship between family background and outcomes.
Recent evidence from the city of Milan, focusing on the third year of lower secondary school, confirm these trends: class composition has an additional effect on learning outcomes beyond the already known influence of individual characteristics (Cordini & Parma, 2025).
If the territorial component is a fundamental dimension of inequality, this paper aims to show—using Milan as a case study—how school segregation is only partially an expression of territorial inequalities. These

inequalities can be exacerbated by mechanisms specific to quasi-market education systems, which can be synthesized in parental choice. However, these mechanisms develop in specific contexts where the local dimension plays a central role in shaping geographies of opportunity and exclusion, by expanding the range of choices for some families and limiting them for others. And while family characteristics play a predominant role in these dynamics, territorial characteristics are equally important, especially in defining the *educational landscape* (Boterman, 2019) within which families operate.

This study uses data from the School Registry of the Municipality of Milan (ANASCO) to reconstruct the intensity of the phenomenon in the first educational cycle, using one of Italy's main cities—characterized by strong migratory flows—as an example (ISMU, 2024). This database includes information on every resident of compulsory school age and each student enrolled in a public or private school within the municipal territory. For every individual, it documents their enrolled school and the school catchment area of residence (or municipality of residence for non-residents in Milan). Additionally, it contains various individual-level information such as gender, year and country of birth, citizenship and grade level. For the 2018/2019 school year, socioeconomic status indicators (ISEE) for students enrolled in the school meal service were link to the database.

These geo-referenced data make it possible to reconstruct mobility trajectories within the Milanese school system and to answer the following questions:

1. Which territories show the highest levels of outbound mobility?
2. Which schools are the most attractive?
3. What characteristics define territories with high levels of outbound flows?
4. Which student profiles are least likely to enroll in their catchment-area school?

The results show that mobility is widespread in the Milanese school system: in 2023/24, 40.6% of primary school students and 40.3% of lower secondary school students were enrolled in their catchment-area school. Moreover, 21% of Milanese residents attended private schools—with significant differences between Italians and foreign-born students and across socioeconomic levels. Mobility within the public school system is also very strong and includes students with a migratory background and lower socioeconomic profiles. The analysis of mobility trajectories reveals that, for Italian families, catchment areas with high outbound flows are typically peripheral zones characterized by low socioeconomic levels and concentrations of critical issues. The most attractive schools are located both in the city center and in intermediate zones—mobility is widespread but often geographically limited—as well as in certain peripheral schools that show micro-territorial polarizations within the same neighborhood. For foreign and socioeconomically disadvantaged families, mobility geographies are less clear: movements occur not only along the center-periphery axis, but also horizontally among peripheral schools in disadvantaged districts. Analyzing flows against territorial characteristics helps to understand the relationship between the spatial distribution of disadvantage and school composition, highlighting pockets of territorial and educational disadvantage or, conversely, upward segregation (e.g., toward city center schools), and more dynamic territories where flows significantly reshape the school population relative to that of the broader territory. The georeferencing of data also supports policy interventions at both school and territorial levels, emphasizing both dimensions in designing strategies to combat educational inequalities.

Keywords: School choice, white flight, mapping, school segregation

School Segregation, Family Choices, and Mobility Trajectories in Primary Schools in Turin

Ilaria Lievore – Veronica Mobilio

Introduction:
Equity and inclusion are central priorities in European and national educational policy. Within this framework, school segregation—characterized by the high concentration of students with similar socioeconomic or ethnic backgrounds in specific schools beyond a critical threshold—poses significant challenges. When segregation occurs, minority students often cluster in particular schools, deviating from the average student distribution in the same territory. These "segregated" schools frequently host higher proportions of disadvantaged groups, face greater marginalization, and exhibit lower academic performance (Cordini, 2019).
School segregation is a multifaceted phenomenon shaped by numerous social and geographical factors. Indeed, geographical divides, usually conceived and examined at a national level, can affect social and educational inequalities also at the smaller geographical level, such as the urban context. Existent geographical barriers in the urban context in terms of concentration of disadvantaged families in specific areas can determine high levels of segregation, which in turn can exacerbate social inequalities by fostering or reinforcing cultural, social, and economic barriers, reducing opportunities for student interaction and exchange. This dynamic often leads to long-term adverse effects on students' educational outcomes (Owens, 2018).
Existing research has largely focused on the role of schools in shaping or mitigating segregation, exploring strategies for educational institutions to attract a more diverse student body and address social disparities (Cordini, 2019). Institutional frameworks—such as policies granting families significant freedom in school choice—can either reduce or perpetuate segregation. Spatial factors (e.g., urban environment, home-to-school distance) and the socioeconomic composition of schools play a critical role in these dynamics. Schools' capacity to attract specific student demographics is closely tied to how they communicate their values, objectives, and extracurricular opportunities to families, thereby influencing parental school choice (Ranci & Pacchi, 2007). Yet, it remains unclear whether school segregation primarily drives family choices or merely reflects them. Families may deliberately select schools based on their perceptions, preferences, or knowledge of the student population's ethnic and social composition. This issue is particularly relevant in the Italian context, where families enjoy considerable freedom in school choice, unrestricted by residential zones.
Objectives and Research Questions: This study aims at describing the dynamics of school segregation, as the result of geographical divides within an urban context, and mobility trajectories for primary school students. The urban context of focus is the city of Turin (Italy), which is characterized by strong socioeconomic and ethnic divides within its borders. Our aim is: 1) describing school segregation, both socioeconomic and ethnic, and mobility trajectories focusing on different geographical levels and with a longitudinal perspective; 2) exploring the relationship between school segregation and mobility trajectories (family choices) at the individual level, analyzing the role of students' migratory and socioeconomic background, to understand to what extent territorial differences in school composition can
affect students mobility within the city; 3) understanding the relationship between school segregation and students' competences in the medium-term.
Data: To investigate school segregation and mobility trajectories, the analysis utilizes unique data derived from the Municipality of Turin, collected through book vouchers provided to families for free primary school textbooks[1]. This dataset includes detailed information on students' school enrollment choices and their

[1] The current research is the development of a research carried out in 2020-2023 by Fondazione per la Scuola in collaboration with the Municipality of Turin and two departments of the University of Turin: the Department of

residential school catchment areas, linked to administrative records on family socioeconomic and migratory backgrounds. This study covers primary school students enrolled in the first grade in public schools in Turin between the academic years 2016/17 and 2022/23 (around 32,000 students). The analysis is complemented using INVALSI data for the same school population, focusing on the municipality of Turin. Relying on INVALSI data, three panels are constructed following students in grade 2 the academic years 2017/2018, 2018/2019 and 2020/2021 until grade 5 respectively in the academic years 2020/2021, 2021/2022 and 2023/2024. The parallel use of INVALSI data allows the investigation of the relationship between school segregation and students' competences in the medium-term (3 years later).

Methods: The analytical strategy is divided in three phases: the first two rely on book vouchers data, while the third one relies on INVALSI data:

1. **Constructing indicators of ethnic and economic segregation, and of student mobility**. School segregation is constructed as risk ratios comparing the percentages of (i) immigrant students and (ii) students from low socioeconomic backgrounds at the city, district, and school catchment levels. Student mobility is constructed as the percentage of students enrolled in a different school catchment area compared to the one they live in. At the macro level, these indicators are described with a longitudinal perspective focusing on different geographical levels within the city of Turin.
2. **Analysis of the relationship between segregation and student mobility** on the whole sample (from 2016/2017 to 2022/2023) using two separate linear multilevel regression models for socioeconomic and ethnic segregation, with students at level 1 nested in school catchment areas at level 2. The dependent variable is a binary measure of whether a student enrolled in a school outside their residential district or catchment area. Independent variables include ethnic and socioeconomic segregation measures at the school catchment level, with a one-year lag with respect to student mobility, and controlling for students' socioeconomic and migratory backgrounds, and fixed effect for district and academic year.
3. **Analysis of the relationship between school segregation and student achievement**. The aim is investigating how school segregation measured at the school level in grade 2 might affect students' achievement measured through INVALSI tests in literacy and numeracy 3 years later (grade 5)[2].

Results: The findings reveal strong territorial variations in levels of ethnic and socioeconomic segregation. Geographical proximity between home and school appears to influence family choices: 84% of students are enrolled in a school located within their own residential district, while mobility is higher toward neighborhoods characterized by lower levels of segregation. However, when analyzing the level of school catchment areas (SCA), dynamics emerge showing the convergence of students with similar characteristics in specific areas: in highly segregated neighborhoods, SCA with low levels of segregation can be found, and vice versa.

Moreover, short-range mobility is high: 45% of students attend a school in a catchment area different from their residential one. This suggests the need to focus analysis on the smallest geographical scale to better understand the interactions between segregation and mobility. School segregation is confirmed as a significant factor in shaping short-distance mobility patterns, although other factors also come into play, such as school reputation, perceived quality of the educational offering, or practical considerations like proximity to the workplace or relatives.

Considering SCA within neighborhoods, where wide variations in segregation levels are observed, it emerges that catchment areas with a higher presence of disadvantaged students tend to be less attractive to families than those with a more mixed composition. The average differences in mobility amount to 38 and 30 percentage points for socioeconomic segregation, and 15 and 8 points for ethnic segregation. Furthermore, family choices do not appear to be driven solely by socioeconomic characteristics, but rather by the interaction between these characteristics and segregation levels, especially in SCAs with a high concentration of disadvantaged students.

The results offer useful insights for guiding local education policies, particularly regarding the definition of school admission criteria, the territorial distribution of resources, and the promotion of greater equity in

Philosophy and Educational Sciences and the Department of Cultures, Politics, and Society. The research committee was composed by Professors Paola Borgna, Renato Grimaldi, and Rocco Sciarrone. The research was divided in two steps: a pilot study based on focus groups and interviews carried out by Valentina Moiso (researcher at Department of Culture, Politics and Society) and a quantitative analysis conducted by Tania Parisi (researcher at Department of Philosophy and Educational Science).

[1] The third phase of the analysis is currently in progress and will be ready for the conference if the abstract will be accepted.

school choice processes. The analysis confirms the importance of addressing educational disparities at the urban scale, where local policies have a greater capacity to intervene. The ongoing analysis of INVALSI data will allow for a deeper exploration of the relationships between school segregation and learning outcomes, contributing to a more comprehensive understanding of educational inequalities in the medium term.

Keywords: School segregation, mobility trajectories, family choices, inequalities

Investigating social issues through geospatial analysis: the case of educational poverty and accessibility to services in Catania, Italy

Daniela Ghio – Elvira Celardi – Agostino Marottoli - Raffaele Auteri

Background: the contemporary social research environment is characterised by a world of increasing complexity and rapid transformation, a phenomenon that has been exacerbated by the digital and technological revolution. This scenario is characterised by an unprecedented availability of information sources and data, which are often heterogeneous, fragmented and not immediately accessible. This context needs a re-evaluation of research methodologies, encompassing the deconstruction of prevailing assumptions and categories, and the establishment of novel connections between data, methods and knowledge. The challenge lies not only in exploring new tools, but also in challenging simplistic dichotomies and maintaining a constant tension between continuity and change as an epistemological lever.
In this landscape, the analysis of social dynamics at the microterritorial level assumes fundamental strategic importance. It is imperative to comprehend the intricacies of social, economic and demographic dynamics across diverse geographical contexts for effective urban planning, the formulation of public policy and the interpretation of inequalities (Murgante & Scorza, 2023). However, this type of survey often faces fragmentation and a lack of detailed official data. Indeed, only recently, ISTAT has begun the development of statistics aimed at understanding reality and territory (ISTAT, 2024).
The proposed study investigates the issue of educational poverty in the City of Catania by analysing the accessibility of cultural and educational resources from a spatial perspective. The present study is addressed within the broader context of social research that utilises spatial data as constituent elements of the observed issue. The research is initiated from the research question of how much they can influence the issue in question, and how much the micro-territorial component can influence the spread of the issue to neighbouring territories, or reveal possible spatial similarities/clusters in the distribution of educational poverty.
Methods and data: the methods and data employed in this analysis are characterized by a sophisticated integration of technological and methodological processes. These processes encompass the implementation of algorithms designed for address normalization, linkage, geocoding, management of data upload flow, and the application of a data-driven ontology.
In order to effectively and critically use the data, it was necessary to understand the construction of the data itself. These data are not "raw" or "neutral" in the conventional sense, but rather the result of a specific project involving specific processes aimed at refining the information. Additionally, methodological choices with potential limitations, such as linkage quality or geocoding reliability, were involved. In this sense, the study on Catania was characterized in its initial stages by a revision of the geospatial geometries derived from the administrative data to conform them to the residential reality of the municipalities; by a search and normalization of the addresses of the points of interest (e.g., schools) lacking in the sources of geographic coordinates; and by the implementation of a relational database for linkage and integration of information through specific queries. This preliminary activity was essential to establish a database endowed with consistency and informational quality and conforming to the operational logic of the algorithms that insist on geocoding and related metrics by the software tools and services used.
Geospatial analysis provides data at a level of detail that was previously inaccessible with official sources disseminated on a large scale. This newfound level of detail enables new research questions about micro-local issue and necessitates a rethinking of research designs to take full advantage of this granularity. This consideration inspired research on educational poverty in Catania, with the objective of refining theoretical models on social stratification, spatial inequalities, and the impact of spatial context on educational poverty. The research strategy involved the integration of micro-territorial socio-demographic data with geographic and accessibility information. Specifically, the objective was to utilize the concept of spatiality to derive

micro-territorial data, which would be instrumental in assessing potential disparities in access to services by the school-age population. This was achieved by proposing the creation of composite indices of geospatial accessibility and, in parallel, conducting a spatial autocorrelation analysis.
The accessibility indicators developed quantify proximity to cultural services (e.g., schools, libraries, theaters, museums, cinemas) and public transportation coverage (in terms of travel times and distances). The utilization of Geographic Information System (GIS) tools and spatial analysis facilitated the modeling of spatial nodes and their relationships in graphs. This approach enabled the identification of optimal routes and concentration/dispersion phenomena.
To this end, a range of sources were consulted: The data set includes the following elements: Istat census data on population; Istat data on indicators of economic hardship and early exit from the education system at the municipal level; GTFS data for accessibility to public transportation and geospatial calculations based on centroids to measure the distances between cultural/educational places, with assignment by competence to different municipalities (micro-territoriality).
The spatial processing and analysis were facilitated by the utilization of Geographic Information System (GIS) tools, specifically QGIS, in conjunction with the statistical software R. A spatial autocorrelation analysis (Moran and LISA) was employed to discover relationships between adjacent micro-territories. A minimum path calculation between nodes in a graph with nonnegative weights (Dijkstra's and Haversine's algorithm) was subsequently implemented to identify urban public service routes between points of interest. This was achieved by drawing times and distances through queries to OpenStreetMap and OpenRouteService sources via API.
However, the utilization of geospatial analysis techniques and the data derived from these techniques are not without challenges and require a "critical appropriation of the digital" (Leon & Rosen, 2020).
While these techniques exemplify innovative strategies for integrating diverse data, the challenge of obtaining official data at the microterritorial level, particularly census data projected onto fine geography, underscores a substantial information deficit. This deficit limits the depth and robustness of analyses based on administrative or survey data aggregated at larger spatial levels (e.g., municipalities, provinces, or census grids). Given the aforementioned limitations, the spatial analysis model developed is a replicable exemplar for comparing microterritorial spatiality with other social phenomena in different areas, such as migration and urbanization dynamics (Ghio et al., 2022) or counter-urbanization trends arising from internal mobility (Ghio et al., 2024).
Expected results: the Catania study yielded initial results that demonstrated a lack of substantial correlation between the calculated geospatial indices and the ISTAT indicators at the municipal level. Spatial autocorrelation analysis (Moran, LISA) did not reveal significant patterns, suggesting (within the limits of the indicators compared) a random distribution of educational poverty among the municipalities analyzed, given the information available. However, the analysis did reveal an inverse correlation in the area of school dropout by foreign youth, suggesting further investigation of ethnic spatial concentration in Catania.
The current implementation refines the modeling of distances and travel times for multiple means, even combined means (e.g., buses with pedestrian transfer), based on the urban-metropolitan road network and the social "weights" attributed to nodes and spatial relationships between nodes.
Examples of these challenges include traffic congestion, inadequate accessibility infrastructure, divergent perceptions of safety, housing concentration, and related issues.
The subsequent implementation will entail the following: firstly, the replacement of predefined spatial grids with centroids of residential units, with the objective of enhancing data granularity; secondly, specific analyses by socio-demographic groups, including, but not limited to, age groups, gender, migration status, and others.
Contributions: this research experience in Catania, which is currently evolving, reflects the issues of how to navigate in a context of heterogeneous and fragmented data, how to integrate different sources to answer complex questions, and how the availability (or not) of data at a certain level of granularity conditions research questions and designs. It exemplifies a concrete case in which the conventional understanding of "data construction" is confronted by the "massive use of big data" (GTFS in this instance, though the principle remains consistent), and the necessity to integrate information from disparate repositories to overcome the limitations imposed by official sources at a particular scale.
The study explicitly noted the limitations imposed by the lack of official microterritorial data, indicating that the availability of census and specific data on educational poverty at a microterritorial level of detail could lend an "additional level of significance" to the model adopted.

The intended contribution directly addresses the need to "rethink social research" through "innovative strategies of inquiry" in the current context of "overabundance of information sources and data, often heterogeneous and fragmented." The study's primary focus is on the deconstruction of assumptions concerning the accessibility of aggregate data and the necessity of conducting investigations at the micro-territorial level. The study explores the intricacies of integrating data from diverse sources through geospatial analysis techniques and tools, addressing the management of "complex data ecologies." The text goes on to discuss the use of highly automated processes and the importance of "critical digital appropriation" to consciously use the data produced by such infrastructures.
At a broader level, the study emphasizes the need to improve infrastructure for official statistics that integrates the spatial dimension, such as the Basic Statistical Registry of Places (RSBL). The development of such infrastructures is predicated on the expectation that they will achieve quality, detailed, geospatially integrated official statistics. This, in turn, will enable them to overcome microterritorial limitations and facilitate new research that is critical for topics such as sustainability, climate, and disaster risk. These resources offer the potential to enrich theoretical frameworks, support policy and planning with detailed information, and facilitate the translation of research findings into practice.
The advent of such technological developments has rendered geospatial analysis a pivotal element in the transformation of territory from an illustrative element to an explanatory variable. This transformation serves as a criterion and a tool for the integration of disparate information. This finding underscores the necessity for researchers to consider a diverse array of geospatial sources and non-traditional APIs, integrating them critically to comprehensively address research inquiries at the microterritorial level. This approach is particularly salient in light of the intricate and fragmented nature of available data, as well as the substantial time required for the development of official statistical infrastructure.

Keywords: Spatial analysis, service accessibility, educational poverty, territorial divide

SESSION 4. LEARNING AND ACADEMIC ANALYTICS FOR EDUCATIONAL QUALITY: TOOLS FOR PEDAGOGICAL ACTION AND SCHOOL GOVERNANCE

ORGANIZER: INVALSI

COORDINATOR: DONATELLA POLIANDRI

19TH NOVEMBER: 4.30 P.M. - 6.30 P.M. {AULA 4 – RESEARCH 5}

Enhancing School Evaluation and Planning through Learning Analytics – Cases from Europe

Martin Brown – Donatella Poliandri – Cecilia Saint Pierre - Margot Joris

The purpose of this paper is to examine the potential of Learning Analytics (LA) to enhance school evaluation and planning, identifying both the affordances it presents and the constraints inherent in its implementation. Drawing on empirical evidence from case studies conducted in Belgium Flanders, Ireland, Italy, and Spain and as part of the European Commission funded Erasmus+ project (QUALAS+ Quality Assurance and Learning Analytics in Schools), this study employs a cross-national comparative research design of which the sample comprises 20 secondary schools purposefully selected in each country to reflect diverse school types, socio-economic backgrounds, and varying degrees of digital maturity. Within this, data was gathered through semi-structured interviews with key stakeholders, including school principals, deputy principals, digital learning coordinators, teachers with responsibility for evaluation, newly appointed teachers, students, and parent representatives.

Thematic analysis of the qualitative data revealed a complex interplay between school culture, leadership, data literacy, and infrastructural readiness in shaping how LA is perceived and utilised at the school level. While policy frameworks in each country promote data-informed decision-making, in practice, the uptake of LA remains uneven. In schools where collaborative cultures, distributed leadership, and technical supports are present, LA has begun to play a formative role in teaching, learning, and school development planning. On the other hand, low levels of staff capacity in data literacy, and concerns around Artificial Intelligence, data privacy and ethics often inhibit less meaningful engagement with LA tools. Finally, a recurrent theme across all three contexts is the necessity for targeted professional development for school leaders to build data literacy and teacher confidence in leading, interpreting and acting upon LA outputs.

The paper argues that LA can enhance the rigour and reflexivity of school self-evaluation. However, its transformative potential depends upon enabling organisational conditions, ethical safeguards, and sustained investment in the professional capital of educators.

Keywords: Learning Analytics, School Self-Evaluation; Data Literacy, Teacher Agency, School Leadership

A Critical Review of Learning Analytics in Education through Socio-Technical and Situated Lenses

Stefania Capogna

This paper aims to critically explore the transformative potential of *Learning Analytics* (LA) in school settings, questioning the epistemological, organizational, and pedagogical conditions that enable the conscious and reflective use of data for educational purposes. In particular, it assumes a socio-technical and multidimensional perspective, in which the analysis of digital data generated in learning processes and school information systems is read not only as a technical opportunity, but as a discursive and situated practice, capable of re-shaping the relationships between actors, institutions, and knowledge.
The emergence of *Learning Analytics* (LA) as tools to support teaching, personalization of learning, and school governance represents one of the most promising—and also more controversial—guidelines of contemporary educational innovation. While the potential of such tools to improve the quality of education has been widely recognised (Ferguson, 2012; Capogna, 2018; Ifenthaler & Yau, 2020), on the other hand, numerous studies warn against the undesirable effects and the logic of technocratic reductionism and depersonalization of the educational relationship that may accompany their use (Evers &. Kneyber, 2016; Selwyn, 2020; Williamson et al., 2022).
This contribution proposes a literature review to explore the main emerging approaches, tensions, and perspectives in the field of LA. An interdisciplinary theoretical framework guides the analysis, reflecting on the need for a paradigm shift that leads to overcoming the view of data as neutral entities and learning to consider them as discursive and situated practices that participate in the construction of educational meanings and power relations.

Keywords: Learning analitics, school analytics, education, teaching/learning

Students' Outcomes at the State Exam: Analysing Students' Choices in Written Tests to Improve Learning Pathways

Paolo Davoli – Francesco Orlando

Introduction

Each year, approximately 500,000 students in Italy take the *Esame di Stato* (State Exam) at the end of upper secondary education (in 2025, 524,415 candidates, including 13,066 external candidates).
Avoiding unnecessary technicalities, the general structure of the exam is summarized by MIM (2025a); we highlight here certain aspects relevant to this contribution. Students are admitted to the final exam even with a single failing grade and can accumulate up to 40 credit points out of 100, awarded by the class councils of the third, fourth, and fifth year, based on grade point averages and other parameters (including, from 2025, conduct).
The first test, written composition in Italian requires, students to choose one of seven prompts issued by the Ministry. The second written test concerns one or more subjects specific to the student's track, defined annually by ministerial decree. For technical and academic high schools (licei), the test is prepared centrally by the Ministry; for vocational schools (professionali), it is prepared by the individual examination boards based on a national framework.
Each of the three exam components—the first and second written exams, and the interdisciplinary oral interview—is scored out of 20 points. Out of the total 100 points, 40% is derived from school-based assessment (credits), and 60% from final examination performance. Under regulated conditions, examination boards may award up to 5 bonus points and bestow honors on students who achieve the maximum score without bonuses.

Research Objectives and Hypotheses

For each candidate, the Ministry's information system integrates multiple data sources through the National Student Registry, including educational trajectories, school administrative data, and information entered during exams via the *Commissione Web* platform. These are linked with background variables such as gender, nationality, regional location, academic delay or advancement, and school credit. They are further associated with exam-related administrative-didactic data, such as scores for the first and second written tests, bonus/honors awards, and the students' choices during both written tests. The result is a substantial *Academic Analytics* dataset.
This study focuses specifically on students' choices in the second written test. For technical and academic tracks, national guidelines (Ministerial Decree 769/2018) define the structure and content of the second

exam, listing core topics for each subject and outlining evaluation rubrics, which examination boards adapt with level descriptors.
At science-oriented high schools (licei scientifici), the second written test involves solving one of two proposed problems and answering four out of eight proposed open short-answer questions. In technical institutes, students complete a mandatory first section and choose two out of four questions in the second section. Although such choices might appear didactically irrelevant, this is not the case. In science-oriented schools, for instance, math problems tend to concentrate on well-covered areas like functions or calculus, whereas the eight questions may touch on underrepresented topics (e.g., probability and statistics). Similar dynamics are observed in technical institutes.
Exploring how test choices relate to variables like gender, nationality, geographic region, and individual academic history can yield valuable insights into the actual teaching practices in schools, student challenges, and their learning trajectories.
While the Ministry annually publishes summary data on exam outcomes, e.g. candidate preferences in the first test (MIM 2025b); or regional statistics on honors and final scores (MIM 2024), but these are aggregate figures with no correlation analysis.
At the regional level, prior studies (Davoli, 2016; Frolloni & Davoli, 2023) have analyzed detailed ministry data to correlate student background, school pathways, exam results, and individual choices. Further studies explored specific topics, such as outcomes from four-year programs (Davoli & Olivetti, 2023; Davoli & Orlando, 2025).
To the best of our knoledge, no academic studies based on national-level persistent datasets have analytically examined the relationship between subjective variables (e.g., students' choices of exam prompts), objective background characteristics (e.g., gender, nationality, educational track), and objective outcome variables (exam results, school credits). This study aims to address that gap, focusing on students' choices in the second written exam in science-oriented licei and in the main technical tracks, where these decisions are most easily observed.

Data and Methodology
Using the national dataset to be released by the Ministry in late July 2025, the analysis will focus on science licei and the most widespread tracks within technological institutes. These data types were selected because they are now available for the first time and for the supposed easiness of connection between methodological content of the question and the student choices in the second test. The dataset comprises two cohorts of approximately 100,000 records each.
The analysis will examine relationships with variables such as gender, nationality, macro-region, academic delay/advancement, and pre-exam academic performance (credits). Descriptive statistics and, where possible, multivariate analysis techniques will be employed. Historical data, such as bonus and honors awards, will also be included.
Findings may support both ministerial teams designing exam prompts and schoolteachers, offering insights to enhance teaching practices in core disciplinary areas.
Results
Analytical data on second exams will be available in July 2025. Preliminary findings from earlier regional-level analyses (Emilia-Romagna) by Davoli (2016) and Frolloni & Davoli (2023) on other variables include:

- Educational track choices by geographic origin. Almost half of regional students choose academic high schools (licei). Students from other Western countries, as well as from China and the Middle East, show similar patterns. Students from Eastern Europe (EU and non-EU) show an almost half of the preference for licei. For students from other countries, enrollment in academic high schools tends to decline, with a significant shift toward vocational institutes.
- Admission rates by school type: In the region analyzed, 3.4% of students are not admitted to the final exam; this drops to 0.9% for licei (1–2 students per school) and rises to 6.6% for vocational institutes (over 1 per class). These patterns are stable over time and gender-biased: males are more than twice as likely not to be admitted as females. Similar trends are observed in lower secondary school exams. Among students not admitted, nearly 24% are non-Italians, though they represent less than 10% of the total—highlighting the need to address these selective dropouts.
- First exam prompt choices of students in 2015 clearly show correlations with students' school credits. Higher-performing students prefer analysis of a literary passage and artistic themes, while those with lower credits prefer general or technical-scientific topics. This polarization may reflect differences in students' writing skills, allowing each student to gravitate toward the prompt that

best matches their strengths. Nonetheless, it also risks reinforcing divisions and hierarchies in students' academic trajectories.

- Over 99.5% of admitted candidates pass the exam. However, diploma score distributions differ significantly by school type, both in the examined region and at a national level. Fewer than 20% of licei students score between 60–70, while nearly 35% of vocational students fall in this range. The opposite occurs for scores between 91–100. Female students are overrepresented in the higher score ranges, while males dominate lower ones. Non-Italian students are overrepresented in lower score bands. Around 25% of Italian students receive final exam scores between 60 and 70, while this proportion rises to over 35% among students without Italian citizenship. In contrast, when looking at the highest score band (91–100), 25% of Italian students achieve this level, compared to just 16% of non-Italian students. These disparities point to persistent outcome gaps and highlight the need for targeted measures to address systemic inequities in student achievement.
- 32.7% of graduates received a bonus. Among licei students, the rate was 38%; among vocational students, 24%, with technical institutes in between. One might say that advantage begets advantage: the bonus tends to enhance already higher average scores among students from academic high schools (licei) compared to those from vocational institutes. This, too, raises important questions about the overall fairness of the educational system and its final assessment procedures. Notably, the bonus was not awarded to all students who met the eligibility criteria, indicating that the examination boards exercised discretion rather than applying automatic rules, taking into account the specific circumstances of individual candidates and their class contexts.
- In the examined region, 2.6% of graduates received honors, two-thirds of whom were female. Among students eligible for honors were 3.4%, again confirming boards' evaluative autonomy. Notably, 66% of all honors were awarded in licei. Statistically, a licei student is nearly four times more likely to earn honors than a vocational student, and nearly twice as likely as a technical student. This points to systemic bias in exam procedures, which appear to favor conceptual and academic performance over technical and applied competencies.

Keywords: Academic analytics, student assessment, educational improvement, school dropout

Making Sense of Data, Rethinking Spaces: A Participatory Reading of School Improvement from the Students' Perspective

Sebastiana Fisicaro

Introduction

The growing use of educational data has encouraged the adoption of learning and academic analytics. This contribution proposes a reflective and contextualized approach, focusing on student voice and learning environments as pedagogical and strategic elements in school self-evaluation processes.

Research Object and Aims

The research focuses on the relationship between data use in school evaluation, listening to students' voices, and the design of school spaces. It follows two directions: (1) qualitative analysis of strategic documents from ten secondary schools in the Province of Syracuse (funded under the PNRR), to understand how physical environments are described, designed, and utilized in improvement processes; (2) a survey exploring students' perceptions of formal and informal education, their participation in evaluation, and their experience of school spaces. The goals are: promoting a critical and integrated reading of data; exploring sustainable

participatory analytics; valuing student subjectivity as a source of knowledge and a driver of school change.

Data Used

The first part examined public documents (PTOF, RAV, PdM, and 'Scuola in Chiaro' profiles) from ten upper secondary schools in the Province of Syracuse. Focus was placed on how physical spaces are planned and described in relation to pedagogical goals and strategies for inclusion and well-being. The second part involved 51 students (aged 14–17, 78% female) using an online mixed-method questionnaire co-designed with a group of student facilitators. The tool included both closed and open-ended questions, allowing for the collection of descriptive statistics and high-density narrative data.

Interpretation and Obstacles

Despite increasing awareness, national and international research highlights several challenges: low levels of data literacy among school staff, poor interoperability among digital platforms, and fragmented practices focused more on formal accountability than meaningful improvement. A shift is emerging that regards data not as mere control instruments but as levers for generating educational agency, organizational awareness, and continuous professional development.

Transformative Analytics

A transformative approach to Learning and Academic Analytics can support a more formative, reflective, and equitable assessment culture. A comparative analysis of 2024 data on 99 adolescents revealed a significant link between cultural consumption, digital technology use, and learning fragility. In the 17–18 age subgroup, 90% were fragile or at risk on standardized tests despite intensive digital use for personal purposes. This indicates that data availability alone is insufficient to promote equity without a critical, pedagogically informed use of analytics.

Systemic Effects

It is also essential to critically analyze the systemic effects of academic analytics on equity and resource allocation. While these tools can reveal needs and guide decisions, they risk being used prescriptively, reinforcing selection mechanisms rather than addressing inequalities. Only school governance that interprets data in light of local contexts and vulnerabilities can turn analytics into a driver of redistributive policies and inclusive planning.

Methodology

The methodology is qualitative and exploratory, based on document-student voice triangulation. Document analysis followed an inductive coding grid across three macro-areas: learning environments, internal evaluation, and well-being. Closed answers were analyzed statistically, and open responses were coded using a grounded theory approach. The theoretical framework draws on Bronfenbrenner's ecological model, the 'school that listens' concept (Rudduck & Flutter), and principles of formative and transformative assessment.

Results

Document analysis shows poor integration between methodological innovation and space design. Environments are often described generically, with little relation to stated practices or learning needs. Some RAVs connect well-being, space, and personalization. Students perceive formal education as fragmented and distant but value informal, collaborative experiences. Over 60% want to be involved in evaluation and space planning. Open responses reference 'empty classrooms' or 'spaces that don't speak,' but also spaces 'that make us feel heard' when well
designed. The data integration suggests a participatory analytics model where student voice acts as both metadata and change engine.

In-Depth Discussion and Policy Implications

Combining document and student perception analysis offers a layered view of the relationship between school and innovation. Schools often remain stuck in a self-referential design logic tied to regulations, with limited dialogue with users. Students call for personalization, flexible and welcoming environments, and emotionally meaningful learning spaces. Over 70% see informal learning (in hallways, peer groups, extracurricular settings) as more useful for developing transversal skills. This underscores the need to recognize hybrid learning as a structural dimension, aligned with EU principles on lifelong and life-wide learning.

Gaps and Opportunities

The PNRR has sparked new investment in learning environments, but risks becoming a mere infrastructural update unless accompanied by shared educational goals. The RAV is underutilized: in most cases, it is a formal requirement rather than a reflective tool. Only two out of ten schools mention student consultation or qualitative monitoring of space impact on learning. Often, methodological innovation (e.g., flipped classroom, cooperative learning) is claimed but not supported by flexible spaces. This gap between declared intent and actual practice undermines coherence. However, where schools engage families and students in needs assessments, perceived quality of environments and evaluation improves significantly.

Conclusion

We need a systemic rethinking of how schools collect, read, and use data. A culture of data as process—dialogic, situated, generative—should replace data as static product. Participatory analytics trigger listening, dialogue, and shared decision-making, giving centrality to all actors, especially students. This shift demands time, care, and a paradigm change in teacher training, governance, and external evaluation.

Literature Review

The findings support the idea that school improvement cannot rely solely on top-down indicators. Research (Rudduck & Flutter, OECD, INDIRE) shows that listening to students is key for understanding needs, motivating learning, and shaping meaningful contexts. 'Pupil voice' initiatives across the UK, Canada, and Nordic countries report positive impacts on engagement and belonging. In Italy, co-evaluation and participatory design are emerging but not yet integral to school culture. Students offer nuanced readings of their contexts, identifying both problems and solutions to renew practices and environments.

Future Prospects

Strategic documents (PTOF, RAV, PdM) should be reformulated as inclusive, narrative tools reflecting the school's identity. Evaluation must capture situated learning, relationships, and transformations. Participatory tools—narrative interviews, spatial perception maps, mixed focus groups—can democratize evaluation. Teacher training should prioritize data literacy and integrated learning space design. European experiences show how analytics, pedagogy, and environment can align within schools-as-learning-communities. The 'Laboratorium' model, as a dynamic ecosystem integrating space, time, practice, and data, offers a scalable solution to today's educational challenges, especially in vulnerable contexts.

Keywords: Learning analytics, Participatory evaluation, School space, Informal education, Student voice

Data Use and Learning Analytics: Teacher Competences and Practices in Five Italian Schools

Donatella Poliandri – Sara Romiti – Grazia Graziosi – Graziana Epifani

Introduction

In recent years, the increasing availability of data in school contexts has fostered critical reflection on the potential of Learning Analytics (LA) to support teaching, personalise learning pathways, and inform school-level decision-making processes. In Italy, academic discourse has gradually expanded to include this topic, exploring the opportunities offered by the systematic use of data in education (Fulantelli & Taibi, 2014; Ranieri, 2015; Bembich, 2019; Agasisti & Bowers, 2021), also in connection with school evaluation and improvement processes (Palmiero & Cecconi, 2019). However, both literature and teacher training experiences show that LA still struggle to enter everyday school practices, hindered by limited data literacy and the absence of structured professional development pathways (Gabbi, 2023; Gala et al., 2019; Morini et al., 2019).

Research focus, objectives and hypotheses

This qualitative study aims to investigate whether and how data and LA are integrated into the daily practices of Italian lower and upper secondary schools, with particular attention to the conditions that enable or constrain their transformative use. The research is based on the hypothesis that the effective use of data does not rely solely on the availability of digital tools, but rather requires investment in teachers' professional competences, an enabling school culture, and structured training and development programmes.

Data used

The study was conducted in five upper secondary schools in the Lazio region, within the framework of the European Erasmus+ QUALAS project, and follows a descriptive multiple case study design (Yin, 2009). The selection of schools was guided by a purposeful sampling strategy aimed at representing a variety of institutional profiles, territorial contexts, and student populations. The sample includes general, technical, and vocational schools located both in urban centres and in peripheral or inner areas, understood as semi-urban or marginal zones in relation to major regional hubs. The selected schools differ in terms of type, size, level of digital maturity, orientation towards innovation, and the socio-economic background of students.

This study is based on qualitative data derived from semi-structured interviews (Corbetta, 1999) conducted in each school, involving both internal and external stakeholders, for a total of 45 participants, including school principals, teachers, students and parents. All interviews were audio-recorded and fully transcribed. The diversity of participants enabled the collection of a broad range of perspectives on the meaning and use of data in everyday teaching, organisational dynamics and school decision-making processes.

Methodology

The methodological approach adopted is based on content analysis as inquiry (Losito, 2007), with the aim of exploring the representations, practices and organisational conditions that characterise data use in schools. Interviews were analysed using an internationally shared thematic matrix, which guided the construction of the qualitative database. Each interview question was treated as an analytical category, while the responses were considered empirical units and summarised following a shared protocol among the researchers.

The analysis followed a comparative logic by thematic area, incorporating selected verbatim quotations to highlight convergences and divergences within and across schools. This abstract focuses on five dimensions: everyday data use, awareness and diffusion of LA in schools, use of LA for teaching and learning, competences and capacities, and professional development opportunities.

Findings

The results reveal significant variability in how data are used in school practices. In some schools, data use is fragmented, delegated to a few individuals, and lacks a systemic perspective; in others, data are integrated into reflective and collegial practices, supported by a favourable organisational culture and distributed leadership.

Some key findings are outlined below:

• Data use and Learning Analytics

Across the five schools, data use to support teaching and organisational practices occurs in diverse forms and for different purposes. All schools show widespread use of data from internal evaluations, systematic observations, school-based assessments, and national standardised tests (INVALSI), often employed to monitor student performance or for accountability. However, only in some cases do these data inform structured and shared reflection within departments or working groups.

Within this broader context, the use of Learning Analytics – understood as the analysis of data from digital learning environments or technological tools – remains limited and mostly confined to local initiatives or specific roles. In some schools, more engaged teachers (e.g. digital facilitators or evaluation coordinators) use dashboards, digital tests or analytic reports to personalise instruction and provide targeted feedback; in other cases, LA are perceived as technical tools external to teaching practices and remain under-integrated.

The most advanced experiences show that, where data – including digital data – are interpreted reflectively and collectively, they can become resources for designing targeted interventions, supporting student self-evaluation, and making teaching more inclusive. However, the integration of LA into school practices remains a work in progress, heavily shaped by professional cultures and the schools' capacity to generate shared meaning around data use.

• Competences and capacities

Interview analysis shows that competences related to data use and LA are unevenly distributed within schools. Data collection, interpretation, and use are often concentrated among a small group of teachers with specific responsibilities (e.g. data coordinators or digital facilitators). These figures may act with a certain degree of autonomy, often serving as intermediaries between data sources and the broader teaching staff. The findings also reveal limited opportunities for knowledge sharing within departments or working groups, and only sporadic attempts to address data use collaboratively or systematically. This situation tends to reduce the opportunities for the wider teaching staff to become familiar with available tools and develop a shared pedagogical interpretation of data.

• Professional development

Training emerges as a critical factor for effective data use in the schools studied. Professional learning opportunities are often limited and fragmented, typically focused on the technical use of digital tools or on reporting, rather than on data interpretation for formative assessment or instructional purposes. Although all schools involved benefitted from recent PNRR funding for digital innovation, the training courses delivered have mainly addressed new educational technologies and AI for teaching. Rarely are training paths designed in response to self-evaluation outcomes or aligned with the strategic objectives outlined in the school's Development Plan (PdM) and Educational Offer Plan (PTOF).

The schools most capable of activating intentional and reflective uses of data are those that have experimented with integrated, practice-based or cross-departmental training. However, such experiences remain limited.

Conclusions

The transformative use of Learning Analytics in schools requires more than access to digital tools. The evidence collected shows that only when a reflective organisational culture, widespread professional

competences, and coherent, needs-based training processes are in place can data effectively serve as tools for educational improvement.
From this perspective, LA can contribute to the development of a formative, participatory, and equity-oriented evaluation culture.
This study presents certain limitations linked to its qualitative nature and the circumscribed sample: while the findings offer relevant insights, they are not generalisable but rather provide interpretative frameworks useful for understanding local dynamics.
Nevertheless, due to their strong grounding in actual school contexts, these results may inform more conscious training and organisational policies, fostering a data culture centred on professional development and school autonomy.

Keywords: Learning Analytics, Data use, Teacher professional development, Interviews

Exploring the use of Large Language Models for a pre-assessment of Invalsi-like tests by simulating students' behavior

Giovanni Puccetti – Alessandra Boscolo

Introduction

Math tests, and more in general high school tests, have recently found a novel application. Beyond their original goal of testing students' skills and helping them improve, they have been used to test the performance of Large Language Models (LLMs), e.g. ChatGPT. This is done by having an LLM answer the questions similarly to how a student would. The goal is to assess the LLM understanding, and thus its reliability, when answering similar questions and more in general to estimate its reasoning abilities.
In Italian, and specifically using past Invalsi tests, this has been studied in peer-reviewed works [5, 4] as well as in unpublished pre-prints [3]. The authors of [4] show two relevant results providing evidence motivating this work:

- LLMs show a large variability in answering abilities, in particular, very large ones, such as Llama 3.1 405 Billions [2], answer accurately, almost acing Invalsi tests while other, smaller models, perform poorly.
- Models performance is on average comparable to that of students, more so when focusing on Mathematical answers, while when tested on language understanding in Italian models perform better. And there are both models that outperform the average student performance and models that underperform it.

Research Question

Based on these findings about LLMs accuracy on Invalsi tests this work aims to provide early answers to the following research questions:

Research Question

Can we use a variegated set of LLMs, spanning several properties, such as: Generalist – math-focused, English-first – Italian first, Large – Small, etc. To simulate the performance of a class of students? If so, can we use the results obtained by this "synthetic class" to obtain an automated pre-assessment of candidate Invalsi questions?

Data and Methodology

We reevaluate the results shown in [5] by regarding the models tested on the past Invalsi questions as a class of "synthetic students" and using their varying performance as a simulation of the varying performance of real students. The goal is to understand if testing a large number of LLMs with different skills, serves the purpose of validating Invalsi questions, as it is done for actual candidate questions that are tested in pilot studies, before being used in the yearly Invalsi evaluations.

In particular, we use the answers of LLMs by applying to them the measurements used to validate Invalsi tests by the Invalsi institute itself. Specifically, we will use the Rasch model [1] on the answers of LLMs, to have measures comparable to those obtained for students.
The goal is to compare students' and models' behavior more closely thus providing finer comparison than the overall performance reported in [5]. This in-depth analysis lets us understand if the performance of a group of models can be used to simulate the performance of a classroom and therefore be used for a pre-assessment of candidate Invalsi questions.
As a dataset of questions we will use both: existing Invalsi tests, as collected by [5] as well as the outcome of Generative AI and LLMs to possibly generate synthetic questions.

Discussion

The possibility to pre-assess a candidate Invalsi test, that still has to undergo evaluation with students[1], can provide an automated tool for a coarse but cheap way to exclude those tests that are eventually going to be ruled out anyway.

This work provides an early example of a methodology to automatically assess candidate Invalsi tests and it does not intend to automate the creation of these tests, not even as a long-term goal. Instead, it proposes to study the feasibility of a tool meant for a pre-screening and an early ruling out of those questions and tests that clearly are not a good fit for an Invalsi test, but would only be excluded through a more costly and time-consuming evaluations done by measuring the results of students.

Keywords: Large Language Models, Invalsi Evaluation, Generative AI

[1]https://www.invalsiopen.it/prove/come-nascono-le-prove/

SESSION 9. THE ITALIAN UNIVERSITY UNDER PRESSURE: ACADEMIC PERFORMANCE

ORGANIZER: UNIVERSITY OF PIEMONTE ORIENTALE – UNIVERSITY OF MODENA AND REGGIO EMILIA - UNIVERSITY OF BOLOGNA

COORDINATOR: CARMEN AINA - FABRIZIO PATRIARCA - GIUSEPPE PIGNATARO

20TH NOVEMBER: 8.30 A.M. - 10.45 A.M. {AULA MAGNA – RESEARCH 6}

Leniency, and State-Driven Competition: Theory and Evidence from Higher Education

Carmen Aina – Luca Bonacini – Chiara Mussida – Giuseppe Pignataro

This paper refers to the literature on the economics of education and examines how state-regulated competition in higher education provision can influence universities' strategic behavior concerning academic standards. Starting from the hypothesis that institutions compete by adjusting the leniency in granting academic credits, the paper proposes a dynamic strategic game model in which universities respond to competitive changes driven by the Ministry's approval of online courses. In this context, universities face a trade-off between the need to maintain a high reputation and the incentive to retain more students through more lenient standards. The theoretical model predicts that the impact of competition is heterogeneous depending on the initial reputation of the institution: large universities with high reputational capital may respond by increasing leniency, while smaller institutions may tighten their criteria to preserve their position.

The research aims to answer three main questions: whether the increase in regulated competition has led to greater academic leniency, measured by the annual credits earned by students; whether this effect differs based on university size and reputation; and whether the empirical evidence supports the theoretical predictions of the model. To investigate these aspects, the authors use administrative microdata from the National Archive of Students and Graduates (ANS), made available by the Ministry of Universities and Research for the period 2010-2018. The empirical design relies on a difference-in-differences strategy with fixed effects for year of study, course of study, and enrollment year, along with an event study to test the parallel trends assumption.

The results show that the staggered approval of online courses led to an average increase in credits earned (+1.6 ECTS per year), indicating an overall rise in academic leniency. However, the effect is highly heterogeneous: in larger universities, the increase in credits is more pronounced and statistically significant, while smaller universities show more modest effects compared to the larger ones. These empirical findings confirm the theoretical model's predictions regarding the trade-off between reputation and retention.

This study makes a significant contribution to the debate on public policies in the higher education sector, highlighting how the regulated expansion of educational offerings can create distorted incentives in the definition of academic standards, with potential implications for the quality of academic degrees. These findings provide useful insights for evaluating the effectiveness and side effects of reforms in the Italian university system.

Keywords: Higher education, academic performance, leniency, telematic universities

Shaken but not Trembled: Assessing Academic Performance in the aftermath of the Emilia-Romagna Earthquake

Luca Bonacini – Majlinda Joxhe – Fabrizio Patriarca – Giulia Santangelo

This study investigates the impact of a traumatic event on educational outcomes by analyzing the effects of the 2012 Emilia-Romagna earthquake on university students' academic performance. Focusing on students enrolled at the University of Modena and Reggio Emilia (Unimore), the primary higher education institution affected by the seismic event, the authors aim to assess whether and how academic performance is shaped in both the immediate and long-term aftermath of a natural disaster.

Background and Motivation

A growing body of literature examines the effects of large-scale disruptions on human capital accumulation, particularly in educational settings. Much of the recent attention has focused on the COVID-19 pandemic, with widespread evidence of learning losses caused by school closures and the challenges of remote learning (Engzell et al., 2021; Maldonado & De Witte, 2022; Schult et al., 2022; Contini et al., 2021). Earlier studies have also analyzed the consequences of natural disasters like hurricanes and earthquakes on student outcomes. For example, in the case of Hurricane Katrina, evacuees experienced initial declines in academic achievement, but long-term improvements due to reallocation and school reforms (Sacerdote, 2012; Harris & Larsen, 2023). Research on earthquakes such as the 2005 Pakistan event highlights that disruption to education can have lasting negative effects unless mitigated by infrastructure resilience or effective compensation (Andrabi et al., 2023; Shidiqi et al., 2023). In the Italian context, studies on the 2009 L'Aquila earthquake found no effect on first-year university enrollment but did find delays in graduation and increased dropout risk (Cerqua & Di Pietro, 2017; Di Pietro, 2018). Similarly, psychological literature emphasizes that trauma can impair concentration, memory, and academic focus, though some individuals may show resilience or adaptive responses (Kearney, 2008; Masten et al., 2008; Lerner & Keltner, 2001). In this context, the 2012 Emilia-Romagna earthquake provides a compelling case for a natural experiment. Occurring in May 2012 with two major shocks (magnitude 6.1 and 5.8), the earthquake significantly affected the Unimore campuses and their student population, creating ideal conditions for causal inference using geospatial and longitudinal data.

Data and Setting

The study combines two main data sources:

Seismic Data: Using PGA (Peak Ground Acceleration) metrics sourced from the Italian National Institute of Geophysics and Volcanology (INGV), the authors classify geographical areas into five levels of earthquake intensity. The epicenter was near Finale Emilia, in the province of Modena.

Unimoredata: A rich administrative dataset integrating detailed student-level information from 2001 onward. For this study, the sample includes 38,000 students and over 370,000 passed exams between 2008 and 2018. Key variables include student demographics, residence, academic history, and exam grades. The sample is restricted to students aged 18–36 who were in their second year of a Bachelor's or Master's program during the 2011/2012 academic year, ensuring that they were already enrolled and had completed exams before the earthquake.

Students are categorized into treatment and control groups based on the intensity of the earthquake in their residential area. This spatial variation enables the authors to compare outcomes for exposed versus non-exposed students, holding constant the time of treatment.

Methodology

The empirical strategy employs a Difference-in-Differences (DiD) framework in an event study setting. This approach allows the authors to estimate a series of treatment effects at different time points before and after the earthquake, capturing both immediate and long-term impacts. The basic DiD model estimates the change in average grades for treated students (those residing in high-intensity areas) relative to control students (residing in low-intensity areas), before and after the earthquake. The regression controls for student fixed effects (to account for unobservable individual traits such as ability or motivation), exam fixed effects (to control for differences in difficulty across subjects), and time fixed effects (to capture seasonality and other time-related shocks). This flexible specification helps isolate the causal impact of the earthquake on student outcomes (Rambachan & Roth, 2023).

Main Results

The results offer a nuanced picture of trauma and academic resilience:

Immediate Effect (Summer 2012): In the session immediately following the earthquake (Summer 2012), students in the treatment group showed a statistically significant increase in their average grades—approximately +0.2 points on a 30-point scale. This counterintuitive finding suggests a short-term performance boost among affected students.

Long-Term Effects: Despite the initial surge, the positive effect disappears in subsequent exam sessions. Over the medium- to long-term, there are no statistically significant differences in academic outcomes between treated and control students.

Robustness: Alternative specifications confirm the main findings. When controlling only for exam or student characteristics separately, the estimates vary. However, the most rigorous model, including both student and exam fixed effects, produces consistent evidence of a short-term positive effect and no long-term impact.

Sessional Differences: Most exams occurred in the winter (35%) and summer (41%) sessions, with minimal variation by session affecting the core conclusions.

Interpretation and Discussion

The authors interpret the immediate positive effect as a manifestation of academic resilience in the face of adversity. The trauma of the earthquake may have triggered heightened motivation, a collective sense of solidarity, or enhanced support systems - academic, psychological, or social - that temporarily boosted performance. Alternatively, examiners may have been more lenient or students more selective in their exam choices in the aftermath of the shock. Over time, however, the effects appear to normalize. The initial boost fades, and academic performance returns to pre-earthquake levels. This suggests that while short-term resilience is possible, sustained improvements require ongoing support and structural interventions. The study raises important questions about the interplay between trauma and performance, including the potential mediating roles of psychological adaptation, institutional responses, and peer networks. While some students exhibit resilience, others may suffer lasting negative effects not captured in exam scores alone.

Contributions and Implications

This study contributes to the broader literature on disaster resilience, education, and mental health by showing that the academic consequences of trauma are not necessarily negative nor permanent. Its main contributions include: leveraging a high-quality administrative dataset with longitudinal coverage; using geospatial data to construct a quasi-experimental design; and highlighting both the potential for resilience and the limits of short-term adaptation. From a policy perspective, the findings underscore the importance of post-disaster support mechanisms that go beyond immediate relief. While some students may respond positively to adversity in the short run, institutions should remain vigilant for delayed effects on dropout rates, mental health, and learning trajectories.

Conclusion

Natural disasters pose significant challenges to education systems, yet their effects are complex and multifaceted. This study of the 2012 Emilia-Romagna earthquake finds that affected university students experienced a short-term improvement in academic performance, but no significant long-term effects. These results suggest that resilience can emerge even in the face of trauma, but also that such gains may be temporary. Future research should explore the mechanisms behind these dynamics, examine heterogeneity in student responses (e.g., by gender, socio-economic background, or field of study), and assess broader impacts on mental health, labor market outcomes, and long-term educational attainment.

Keywords: Academic performance, Emilia-Romagna earthquake, Event study, Trauma

(De)Efficiency of Italian Universities: a comparative analysis of Public, Private, and Online Institutions

Tommaso Agasisti – Gaetano Francesco Coppeta

Traditionally, Italian universities operated as public entities managed within the state. However, the Higher Education (HE) sector has significantly transformed in recent decades. A series of reforms have introduced market-oriented mechanisms and competition into the public sector, encouraging universities to compete for funding, attract more students, and enhance research output to secure more research grants. Alongside transforming the public sector, a new and less explored element has emerged in the HE landscapes: the coexistence of different institutions offering the same "products": education and research. Italy represents an interesting case in this respect. In addition to the traditional public universities, several private and online learning universities have emerged. The latter, known in Italy as 'telematics', has experienced remarkable growth, with an increase in student numbers of 180 thousand units from 2011/12 to 2021/22, and has become a significant part of the Italian HE system. The picture for the academic year 2021/22 sees 1.6 million students enrolled at state universities (82.2%), 123 thousand enrolled at non-state universities (6.3%), and 224 thousand enrolled at telematic universities (11.5 per cent), up from 2.5% in 2011/12 (ANVUR 2023). Despite being regulated under the same framework, online universities exhibit distinct characteristics, especially regarding student-to-teacher ratios and employment practices. Current international literature has only begun to explore online universities, with some studies focusing on student outcomes (Bettinger 2017). Regarding organizational dynamics (Deming at al. 2015) show that institutions offering online courses tend to charge lower tuition fees, raising the possibility that online learning might be a technological advancement capable of "bending the cost curve" in HE. However, this comes with concerns that the quality of education may need to improve when more content is delivered online. This paper explores this phenomenon within the Italian context, where online learning has grown substantially. Our primary contribution is to comprehensively describe the Italian HE sector, focusing on student, academic, and non-academic staff characteristics and the research output of various institution typologies from 2010 to 2021. Further, we analyze technological differences and efficiency scores across different institution typologies using a Generalized True Random Effect Stochastic Frontier approach (Badunenko and Kumbhakar (2017). The GTRE approach allows us to decompose inefficiency into persistent and transient components. We observe that online institutions carry both a structural inefficiency component, linked to long-term strategic choices, and a transient one, possibly due to adaptation lags or institutional instability.

Our analysis reveals the emergence of a dual model within the HE sector. On the one hand, traditional universities remain structured around a dual mission of teaching and research, supported by permanent academic staff and public funding tied to performance metrics. On the other hand, online universities have developed a model oriented primarily toward delivering teaching at scale. These institutions operate with high student-to-faculty ratios and a largely contingent workforce, often composed of temporary or part-time instructors. While this model significantly reduces fixed labor costs and allows for rapid enrolment growth, it raises important questions regarding long-term sustainability and academic integrity.

The results highlight key organizational patterns. Telematic universities appear structurally less invested in research activities: they publish fewer scientific papers and allocate limited resources to faculty with research responsibilities. Consequently, when assessing output using a broader definition of university performance, one that includes both teaching and research, their efficiency levels decline.

Furthermore, our findings show that when only teaching outcomes are considered, such as the production process of online universities appear as efficient as their traditional counterparts. This suggests that their model is well-adapted to mass instruction, especially in contexts where flexibility and accessibility are

valued. However, as soon as research is introduced into the production function, their performance suffers. This supports the interpretation that their institutional model has opted.

This divergence reflects broader tensions in higher education systems worldwide. As HE institutions face mounting financial pressures and growing demand for flexible learning, online models offer an appealing alternative. Yet, the shift toward scalable teaching solutions must be critically assessed against potential losses in academic depth, research quality, and faculty engagement.

The study emphasizes a broader issue in contemporary HE governance: technological and organizational innovations, while promising in terms of efficiency, must be scrutinized with respect to their impact on quality. Cost reduction alone does not constitute progress if it undermines core academic values. Institutions that deliver education at low cost but forego the knowledge production function risk contributing to the erosion of the university's traditional role as a centre of research and critical inquiry.

In conclusion, the expansion of telematic universities in Italy illustrates both the opportunities and the challenges of differentiated higher education. While these institutions enhance access and flexibility, their performance in research remains limited. Their growing role in the national system makes it essential to understand how institutional structures shape academic outputs and to ensure that the push for efficiency does not come at the cost of quality and equity.

Keywords: Efficiency, Higher Education, Online Universities, Italy

Pathways to excellence: Efficiency and Best Practices in Italian Universities in Research, Collaboration, and Innovation

Anna Rita Dipierro – Angela Rella – Pierluigi Toma

In recent years, Italian universities have undergone a major transformation, aiming to strengthen connections between academia and external actors within the innovation ecosystem (Compagnucci & Spigarelli, 2024). Through the dissemination of knowledge, universities increasingly contribute to societal and economic development while also playing a critical role in workforce training (Maral & Çetin, 2024). In this context, the concept of efficiency is essential (Kallio et al., 2020; Dipierro & De Witte, 2024) because it concerns how inputs are converted into outputs. This capability is especially important in the university setting, where institutions often operate under resource constraints (Agasisti & Johnes, 2015).

Renowned for its rich history and diverse academic traditions, the Italian higher education sector warrants the attention of practitioners who can support policymakers in guiding ongoing university transformations. Recent challenges demand proactive measures from Italian universities to remain aligned with advancements in knowledge, promote collaboration with industry and institutions, and support innovation (Frondizi et al., 2019). However, to the best of the authors' knowledge, empirical research on the efficiency of Italian universities—specifically in the context of research, collaboration, and innovation—remains limited.

According to the systematic literature review by Rella and Vitolla (2024), there has been limited effort to jointly apply the two most widely used techniques for measuring university efficiency: data envelopment analysis (DEA) and stochastic frontier analysis (SFA). Indeed, defining efficiency is a complex task due to the wide range of interpretations proposed by scholars and practitioners over time. In this study, we adopt the definition provided by Charnes et al. (1978), which considers efficiency as the ability to convert inputs into outputs. There are various ways to calculate efficiency scores, but the most widely used approach involves estimating a frontier that identifies optimal values, followed by comparing the efficiency scores derived from input-output combinations (Førsund et al., 1980; Lovell, 1993; Worthington & Dollery, 2000). Following this overview of the efficiency concept, it is essential to examine how it is measured empirically. As noted earlier, methodological approaches are typically classified into two main categories: parametric and nonparametric. While both aim to quantify efficiency, they differ substantially in terms of structure and underlying assumptions.

In this work, we aim to undertake a comparative evaluation of the efficiency of Italian universities, focusing on key dimensions such as research, cooperation, and innovation (Abramo et al., 2020). Specifically, we

apply the most commonly used nonparametric and parametric approaches—DEA and SFA (Agasisti & Dal Bianco, 2006; Tavares et al., 2021)—to highlight their respective theoretical and methodological advantages and limitations. Each method provides distinct benefits and addresses specific shortcomings, making them complementary tools for assessing the efficiency of complex systems such as universities. These tools are essential for evaluating university performance, especially from a managerial perspective committed to fostering knowledge development. Through a comparative lens, this study employs DEA and SFA to offer a comprehensive assessment of Italian university efficiency in relation to research, cooperation, and innovation activities. DEA and SFA are applied to a data set comprising 62 Italian public universities using three different input-output combinations. This allows us to assess efficiency across multiple activity domains. The data set includes information on research publications, academic collaborations, patent applications, and financial resources, providing a multidimensional and detailed view of university efficiency. By adopting this integrated approach, we aim to answer the following empirical research question (RQ):

RQ: To what extent are Italian universities efficient when assessed using DEA and SFA approaches?

By identifying best practices and offering specific insights into areas where universities can improve performance, this study aims to support the ongoing development of Italian universities and strengthen their global competitiveness in research, cooperation, and innovation. Universities today operate in an increasingly challenging environment, often facing significant budget constraints that necessitate a focus on competitiveness. By optimizing resource management and enhancing efficiency, universities can increase both their attractiveness and competitive positioning (Dipierro & De Witte, 2024). Effective resource management also improves the ability of universities to meet the needs of internal and external stakeholders.

In pursuing these objectives, this study contributes to the academic field by recommending actionable strategies that can enhance the impact of universities on innovation and knowledge production in Italy, thereby supporting broader socio-economic development. The proposed perspective is also adaptable to other European contexts. Specifically, the findings offer policy-relevant insights for resource allocation and institutional reforms within the higher education sector.

Specifically, from a policy perspective, our results emphasize the importance of implementing both nonparametric and parametric tools for efficiency measurement in universities, taking into account the aspects discussed earlier. This is equally significant for university administrators because pursuing efficiency can support the timely achievement of institutional goals. One relevant example is the Sustainable Development Goal (SDG) related to education (SDG No. 4). By managing resources effectively, universities can enhance the quality of education and improve equitable access. Based on our analysis, we suggest that policymakers prioritize investment in research and development (R&D) because it can drive innovation, strengthen technological capabilities, and foster economic growth. Moreover, they should establish mechanisms to evaluate the impact of academic research on the economy, society, and innovation. Regular assessments of funded program outcomes are crucial, with policies adjusted accordingly based on collected data. Policymakers should also introduce regulations that support lifelong learning and continuous skills development. Promoting dialogue and collaboration among public officials, academic institutions, the private sector, and the broader community is essential to align policy objectives with stakeholder needs. Additionally, ethical standards in research—particularly in fields such as data science, artificial intelligence, and biotechnology—must be safeguarded through appropriate legal and regulatory frameworks. Financial systems should be developed to reward universities for efficiency in research output, global partnerships, and innovation. At the same time, institutions should be encouraged to streamline administrative processes and use resources more effectively.

Academics should aim to optimize resource use to enhance both the quantity and quality of research output. They should leverage international partnerships, exchange programs, and collaborations to promote diverse research perspectives and foster global networks of knowledge exchange. Furthermore, fostering inclusive academic environments, offering structured support, and encouraging diverse viewpoints are essential actions. To address complex societal challenges, universities should create systems that promote interdisciplinary and interdepartmental collaboration. Lastly, broader indicators of success (e.g., instructional efficiency, student outcomes, societal impact, and community engagement) should be considered. This comprehensive approach allows for a more holistic understanding of university efficiency.

Keywords: University, Efficiency, Competition, Italian context.

Open the Floodgates or Skim the Cream? Selective vs. open enrollment policies and the race for talent in Italy

Enrico Lippo – Massimiliano Bratti – Daniel Kreisman

Introduction

In recent years, the Italian university system has undergone significant changes in the regulation of admission to degree programs. Within a framework of increasing institutional autonomy, many universities have introduced so-called *numero programmato locale* (NPL), i.e., locally determined selective admission procedures established at the course level. While some fields, such as Medicine, remain subject to national regulation, many other programs have independently introduced access restrictions—often justified by infrastructural capacity, faculty availability, or the desire to improve teaching quality. However, this uneven spread of selective policies across regions has raised concerns about equity, territorial mobility, and the distribution of educational opportunities. Some institutions, typically located in large urban centers and in Northern Italy, have implemented selection in highly attractive programs, thereby limiting access for students from peripheral or disadvantaged areas. Conversely, smaller universities, often located in the South, have maintained or reinstated open access, raising the risk of a two-speed higher education system.

The international literature has widely examined the effects of selective policies. Bleemer and Mehta (2023) show that GPA requirements for access to certain majors in the U.S. disproportionately penalize underrepresented minorities (URM). In Italy, studies such as Carrieri et al. (2015) and Francesconi et al. (2011) have mostly focused on academic performance, reporting mixed effects: in some cases, selection reduced dropout and improved student outcomes; in others, no clear benefits were found. However, there is still a lack of systematic national-level analyses that take into account regional heterogeneity and distributional implications. Our work aims to fill this gap in the literature.

Research questions and hypotheses

This study aims to assess how the introduction of selective admission policies influences student behavior, enrollment patterns, and academic performance. In particular, we aim to answer four main research questions: (1) Which programs decide to adopt selective admission, and what are their characteristics? (2) How does the composition of enrolled students change (in terms of gender, ability, educational and geographical background)? (3) What are the effects on academic outcomes (credits earned, dropout, on-time graduation, final grade)? (4) Who are the students that—despite having a high *ex ante* probability of enrollment—are no longer able to access a program that has become selective?

We hypothesize that selective admission improves average academic performance due to smaller cohort sizes, higher entry quality (cream skimming), and stronger peer effects. At the same time, we expect adverse distributional effects penalizing students with lower educational capital and those from disadvantaged regions. We also expect heterogeneous effects by gender, macro-area, and type of upper secondary school.

Data

We use microdata from the Italian National Student Archive (ANS), covering the entire universe of undergraduate enrollments from 2010/11 to 2022/23. The dataset includes information on degree program (type, institution, admission mode, academic year), students (age, gender, secondary school type, final grade, home province), and academic outcomes (first-year credits, dropout, graduation, final grade).

We reconstruct a longitudinal history of admission modes for each degree program and classify year-by-year transitions from open to selective access. We also match student information with territorial indicators and implement a predictive model to estimate the probability of enrollment in the absence of access restrictions.

In the future, we plan to link this data with INVALSI student achievement records from pre-university schooling, which would allow for a more precise measurement of prior academic ability and better insights into access inequalities.

Method

We adopt a causal approach based on a staggered Difference-in-Differences design (Callaway & Sant'Anna, 2021). The staggered timing of selective admission across similar programs allows us to identify average treatment effects (ATT) by period. We control for program and student characteristics, including fixed effects for degree type and cohort.

The analysis is carried out at two levels: (1) Program level – we examine the impact of selection on cohort size, socio-geographic composition, and average academic performance; (2) Individual level – we estimate conditional models to separate compositional effects from actual treatment effects.
In addition, we implement a Random Forest model to predict each student's probability of enrolling in a given program based on observable pre-admission characteristics. The model is trained on pre-policy cohorts and applied to subsequent ones to identify students with high predicted probabilities who did not enroll after selection was introduced. These "likely excluded" students represent a relevant group for future counterfactual analysis of access equity.

Results and implications

We find that the introduction of selective admission significantly reduces cohort size (–50% on average) but improves the composition of enrolled students: the share of high school graduates from academic-track schools increases (+4.3 percentage points), the average final grade rises (+2.2 points), and local enrollment becomes more prevalent.
In terms of outcomes, we observe substantial improvements: first-year dropout decreases (–3.7 percentage points), on-time graduation increases (+7.8 points), and final grades rise (+0.8 points). However, these gains are not evenly distributed. Conditional estimates suggest that improvements are greater among students with stronger educational backgrounds, while students from vocational and technical schools are more likely to be excluded.
We have planned a future extension of this work to identify "likely excluded" students through predictive models. The idea is to estimate counterfactual enrollment probabilities based on pre-policy characteristics and apply this to post-policy cohorts. We expect that such students will tend to have weaker school performance and come from disadvantaged backgrounds. The findings from this analysis would help assess the welfare implications of selective admission policies and support potential policy responses (e.g., counseling services, targeted support, expanded educational supply).

Keywords: College, selective admission, enrolment, student quality

High-Achieving Peers and Student Performance in Selective and Non-Selective University Programs: Evidence from Italy

Massimiliano Bratti – Enrico Lippo – Andrea Lopes

Introduction

Whether or not peers affect student performance is a long-standing question in economics. Although many contributions have investigated the effect of average peer quality, recently, scholars have focused on the impact of so-called high-achieving (HA) peers (Angrist and Lang, 2002; Arcidiacono & Nicholson, 2005; Feld & Zölitz, 2022; Lépine & Estevan, 2021; Manski, 1993; Modena et al., 2022; Pagani & Pica, 2021). We contribute to this stream of literature by studying the effects of HA students on their peers on several university outcomes in the context of the Italian university system. Italian universities represent an interesting case study because degree programs may differ by level of selectivity at entry even within the same institution or field. This enables us to shed light on the interplay between peer group effects and degree programs' selectivity. Our paper is related to different strands of the literature on peer effects. First, building on the pioneering work of Sacerdote (2001), it contributes to shedding light on the dynamics of peer effects at university level, particularly in undergraduate courses (e.g., Hill, 2017; Feld & Zölitz, 2022). Typically, surveys or data from specific universities have been employed to investigate peer effects in tertiary education, with potential issues of external validity. Instead, we take advantage of individual-level administrative data on the whole population of undergraduate students enrolled in Italian universities, spanning the period from 2010 to 2019. Working with the entire student population, as in our case, enables us to compute precise estimates on peer effects, increase the generalizability of our results, and consider several potential sources of effect heterogeneity, including by gender and degree programs selectivity. In this work, we are also able to contribute to the literature on gendered effects of HA peers, which is particularly relevant for STEM degrees. Mouganie & Wang (2020) explore the impact of HP female peers on female students' choices in STEM fields, using administrative data from China. They find that exposure to a higher proportion of high-performing (HP) female peers increases the likelihood that women will choose a science track in high school, and also has positive effects on college outcomes. On the other hand, they show

that exposure to more HP males decreases the likelihood that women choose a science track and that men's academic choices are generally less affected by peers compared to women's. Our work is closely related to Modena et al. (2022), who investigated the gender-asymmetric effects of HA peers on student performance on the Italian undergraduate student population in 2006–2014. Their results indicate that the positive influence of high-achieving peers is more pronounced among students of the same gender, especially among females. Notably, only HP female students have a beneficial effect on their male counterparts. In contrast, exposure to HP male students can adversely affect low-ability students. We add to Modena et al. (2022) in several respects. First, our data provide course program identifiers, and, accordingly, we can define more precisely a student's peers, leveraging also on the university municipality in which a course is held for degree programs with multiple teaching locations. Second, we investigate peer group effects in more vs. less selective environments, proxied by selective-access vs. non-selective-access degree programs in Italian higher education. Third, we extend the set of educational outcomes to be investigated, including also the probability of continuing in post-graduate degrees, and of choosing STEM post-graduate degrees. Finally, our analysis refers to a more recent period, i.e., 2010-2019. Our paper is also tangentially related to works on rank effects, such as, e.g., Bertoni & Nisticò (2023), Comi et al. (2021), and Carneiro et al. (2025).

Objective

The study investigates the effects of HA peers—defined by the final high school grades obtained from the diploma—on university students' short- and long-term outcomes, including credits earned, dropout rates, and final degree marks, in Italian undergraduate degree programs. The analysis is guided by the following hypotheses:: (*i*) Exposure to HA peers (defined by high school grades greater or equal than 95/100) improves both short- and long-term academic outcomes.; (*ii*) Effects are stronger in selective access degree programs (local/national access modes) due to smaller cohorts and higher peer quality.; (*iii*) HA peer effects are gendered and possibly asymmetric, consistent with findings by Modena et al. (2022); (*iv*) Peer effects do not significantly influence the probability of dropping out.

Data

We leverage very rich and granular administrative data 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 all undergraduate students enrolled between 2010 and 2019 (~2.45 million observations) in all Italian universities. The final dataset used for the analyses includes more than 2.45 million students enrolled in bachelor's or single-cycle degrees between 2010 and 2019, excluding online universities and non-standard programs. The dataset also includes student demographic characteristics (e.g., gender, age, foreign status, high school track, region of origin, etc.), university and degree program identifiers, including information on the municipality of instruction, high school identifiers, and academic outcomes (e.g., credits, dropout status, final marks). HA peers are defined as students with a final high school grade greater or equal than 95/100 (15.50% in our final cleaned sample). The granularity of the data allows precise peer group identification at the degree program-municipality-cohort level.

Methodology

The identification strategy builds on Hoxby (2000) and leverages quasi-random within-degree-program variation in HA peer shares across cohorts. This assumption rests on the idea that while students choose degree programs, they cannot predict their cohort's composition. HA students are identified via their final high school grade. The share of HA peers for student *i* is calculated as the proportion of HA students in the same degree program, municipality, and cohort, excluding *i* (leave-one-out mean to overcome the *reflection problem*, as identified by Manski, 1993). The baseline specification is:

$$y_{ijmdut} = \alpha + \beta_1 \overline{HAPR}_{-i,jmdut} + \beta_2 Local_{jmdut} + \beta_3 National_{jmdut} + \theta \mathrm{X}'_{ijmdut} + \mu \Psi'_{jmdut} + \gamma_{dt} + \delta_{ut} + \varphi_j + \varepsilon_{ijmdut}$$

Which is estimated first on the entire sample and then on each different access modality: free access (non-selective), local test (selective at university level), and national test (centrally managed selective access). y_{ijmdut} denotes our outcomes of interest, *i* denotes the student, *m* the municipality in which the university u provides the degree program *j, d* is the degree class, and t is the student cohort. β_1 is our parameter of interest; $Local_{jmdut}$ and $National_{jmdut}$ are dummies for local and national access types and are restricted to be equal to zero in sample split for different access modalities; X'_{ijmdut} is a vector of student characteristics (HA status, age, foreign, gender, out-of-site, diploma type, region of residence, LP, i.e., Low-Performing student); Ψ'_{jmdut} is a vector of cohort characteristics (log-cohort size, shares of: females,

students with either a scientific or classical diploma, out-of-site students by province and region, foreign students); γ_{dt}, δ_{ut}, and φ_j are, respectively, degree class/year, university/year, and university course fixed effects; ε_{ijmdut} is the error term clustered at the course level.

Results and discussion

HA students represent about 15.5% of the sample, with higher concentration in selective-access programs. Female students are more likely to be HA (17.3% vs 13.2% for males). Dropout rates and performance indicators vary systematically across access modalities. For the short-term outcomes considered (i.e., credits earned in the first academic and calendar year and dropouts), we find that a 10-percentage point increase in HA peers ratio increases first-year academic credits by 0.32 and calendar year credits by 0.35 (equivalent to approximately 3 extra credits over the full range of the ratio). Moreover, we find no significant effect for free access programs, but the HA peers ratio exhibits strong and significant effects in selective programs, especially those with local access. Furthermore, as expected, the logarithm of cohort size is negatively associated with performance, particularly in national-access degrees, confirming that larger classes have a negative impact on students' academic outcomes. As for the long-term outcomes considered (in our case, the final degree mark), we find that a 10-percentage point increase in HA peers ratio increases the final degree mark by 0.20 points (out of 110), which is approximately equal to 2 points in the final grade if we saturate the class with HA student peers. Again, no significant effect is found in free-access programs, while the impact is substantial and significant in selective programs. Being a HA student is associated with a 6.2-point increase in the final degree mark, while low-performing students score 5.2 points lower on average. Also, female students outperform male students by about 0.7 points on average. Finally, the HA peers ratio has no significant effect on dropout rates, which is mainly explained by one's own performance, foreign status, and age. Similarly, HA students are significantly less likely to drop out.

The absence of peer effects in non-selective degree programs is likely due to low concentration of HA students and larger cohort sizes, which may dilute the influence of peers. In contrast, selective programs provide a more concentrated peer environment where HA peer effects can materialize, inducing positive spillovers on peers. Interestingly, local access programs display larger effects than national-access ones, possibly since high school grades are used in selection, making peers more homogeneous in ability. National tests, in contrast, rely solely on standardized exams. The evidence supports the hypothesis that class size and program selectivity moderate peer effects.

Keywords: Peer Effects, Higher Education, Academic Performance, Selective Admission

Early prediction of students at risk of failure: a Machine Learning Approach

Fiammetta Noccioli – Michele Marsili – Marcello Napoli

Introduction

In Italy the phenomenon of university dropout continues to be significant and for that understanding university dropout rates has become a major concern for higher education institutions (Ragni et al.,2024). Alongside this, in 2023 61.3% of first-level (bachelor's) graduates completed their degrees on time, but for the first time in 12 years, study regularity, which measures the ability to complete a degree program within the timeframe set by the curriculum, showed a slight decline (ALMALAUREA 2023).

In response to these challenges, one of the main trends in recent years has been the growing use of predictive models aimed at detecting and preventing academic failure, often focusing on aspects related to students' career and performance. In particular, machine learning techniques are increasingly being integrated into this domain. For example, some Italian studies (Priulla et al., 2024) explored the influence of high-school students' background and performance on universities enrolment choices with a machine learning approach. Other studies (Cannistrà et al., 2024) proved the validity of an innovative multilevel machine learning approach by applying it to Italian university case.

Several international studies have also investigated student failure and dropout prediction at university using machine learning. Some studies focused on comparing individual classifiers and ensemble methods revealed that balanced datasets yield better results (Raj & Vannan, 2020), while others demonstrated that ensemble algorithms like Random Forest and Gradient Boosting outperform Naive Bayes due to its strong feature independence assumption and their hyperparameter tuning considerably improves the outcomes (Keser & Aghalarova, 2022). Similarly, various decision tree algorithms were evaluated for predicting on-

time graduation (Wicaksono & Supianto, 2018), while other researchers have applied diverse evaluation metrics to identify non-active students.

Research object and hypothesis**.** The aim of this study is to predict the risk of student unsuccess defined as either not graduating on time or withdrawing from university, based on data from the first year of university, contextual information, and prior educational background. Specifically, the goal is twofold: on the one hand, to identify the most suitable predictive model between different machine learning models and on the other to understand which predictors are most relevant in this context dealing with the big variation of dropout phenomenon between degrees and universities. For the target construction, in contrast to other studies, students who do not graduate on time were also included among those at risk at university. We believe that this group nevertheless represents a relevant area of interest for universities and policy makers.

Data

This study is based on a dataset created on the combination of different data sources: Ministry of Education and Merit (MIM), University Student Register and INVALSI. The analysis is focused on a cohort of students enrolled in one of the Italian universities in the 2019/20 academic year at a bachelor's degree, which did INVALSI test of grade 13 in the 2018/19 school year and that are present in the National Student Registry, with the exclusion of students without information on credits and academic status. The cohort was followed allowing the observation of the students' academic progress and graduation outcomes. Different types of information are included like individual characteristics, school type, university information, for example number of credits earned at first year, that is a key feature for dropout risk, and changes of study track. The target variable is defined as binary, where 1 denotes a non-completion or withdrawing from university and 0 represents graduation within the standard duration of three academic years.

Methods

The aim is to compare several machine learning models, that have been shown to perform particularly well with this type of data and for classification problems. Specifically, after the pre-processing phase, ensemble methods like bagging and boosting were tested to enhance classifier performance and improve the overall predictive process. Given that boosting models demonstrated better performance compared to other machine learning models, the application of this class was further explored in greater depth. For example, implementations included Gradient Boosting Machine (GBM), Extreme Gradient Boosting (XGBoost) known for its speed and accuracy and CatBoost that handles categorical features natively and efficiently. Then, their specific features and performance characteristics were compared. To ensure robust model evaluation and a better performance, a cross-validation method and a percentage split were used. In particular, a 10-fold cross-validation, that is very popular, was applied by splitting the dataset into ten subsets and a training-test set partitioning of 80% - 20%.

Results

Machine learning offers a powerful approach for predicting university performance and can support targeted guidance and policy planning in the educational sector. The integration of different data sources could be a valuable support in this context to better understand key factors on the unsuccessful university path. The study also confirms the significant role of number of credits at first academic year in students' performance and outcome.

Keywords: Machine learning, prediction models, students' performance, university

STEM at a Distance: Do Telematic Universities Really Narrow the Gender Gap?

Giuseppe Pignataro – Carmen Aina – Luca Bonacini - Chiara Mussida

This study examines gender disparities in Science, Technology, Engineering, and Mathematics (STEM) degree programs, with a specific focus on the role of telematic universities in Italy. Building on recent research suggesting that expanding access to STEM education can reduce gender gaps, the analysis investigates whether telematic institutions--by removing geographic barriers--enhance female participation in STEM.

Using rich administrative microdata from the National Archive of Students and Graduates (ANS) provided by the Italian Ministry of University and Research, the study covers a nationally representative cohort of students enrolled between 2010/11 and 2019/20. The dataset includes detailed individual-level information on secondary school background, academic performance, and higher education trajectories.

Descriptive evidence confirms that although women outperform men in overall tertiary graduation rates, they remain significantly underrepresented in STEM fields. The analysis highlights systematic differences between traditional and telematic university students: the latter tend to have lower academic preparedness, are less likely to hold a lyceum diploma, and are more frequently male.

Multivariate regressions and robustness checks restricted to STEM courses offered in both formats reveal four key results. First, while female enrollment in STEM is lower overall, it is relatively higher at telematic universities. Second, dropout rates are lower in telematic settings, but female students are more likely to drop out compared to male peers. Third, switching between programs is less common in telematic institutions, and the gender gap in switching narrows. Fourth, the advantage women display in on-time graduation in traditional universities is diminished--or reversed--in telematic environments.

Overall, the findings suggest a complex interplay between gender, institutional format, and academic outcomes. While telematic universities provide flexibility and expanded access, they do not eliminate existing inequalities and may pose new challenges related to student persistence and degree completion. Future work will explore heterogeneity across disciplines and use decomposition methods (Blinder–Oaxaca, Gelbach) to distinguish compositional from structural effects.

Keywords: Gender disparities , Telematic universities, STEM degree programs, Academic outcomes

SESSION 11. STUDENTS WITH A MIGRATORY BACKGROUND, MOBILITY AND EDUCATIONAL PATHWAYS: NEW CHALLENGES FOR EQUITY AND INCLUSION

ORGANIZER: INSTITUTE FOR MIGRATION RESEARCH, UNIVERSIDAD DE GRANADA – INAPP - UNIVERSITY OF VERONA

COORDINATOR: PATRIZIA RINALDI - GIOVANNA FILOSA - EMANUELA GAMBERONI

20TH NOVEMBER: 8.30 A.M. – 10.45 A.M. {AULA 5 – RESEARCH 7}

Early Skills Assessment and the Evolution of Educational Inequalities: Evidence from a Longitudinal Study in Milan

Gian Paolo Barbetta – Giulia Assirelli – Alessandro Antonietti – Paola Iannello

Assessing children's skills and competences at an early age is a powerful instrument to understanding and addressing educational inequalities. In fact, previous research shows that learning differences in the first years of primary school tend to persist or even widen over time, with strong implications for future academic performance (Heckman, 2006; Kulic et al., 2019; Duncan et al 2023). In particular, social origin and migratory background are persistent predictors of these inequalities (Bukodi and Goldthorpe, 2013; Azzolini et al. 2019). Also in Italy, recent research shows a strong association between social and migratory background and educational outcomes (Barbetta et al. 2023; Fadda et al. 2023).

This study investigates whether assessing children's skills before the beginning of compulsory education can be informative of their development over time and of the related inequalities. Specifically, we investigate:

1 the existence of early inequalities in children's competences,
2 the predictive power of early skills for later academic outcomes, as measured in grade 2 of primary school,
3 the evolution of educational inequalities between pre-primary and primary school.

To this aim, we exploit a unique dataset that collected a wide set of information on a sample of about 150 children attending 10 pre-primary schools in the city of Milan. We administered the SAI test (Skill Assesment per l'Infanzia, Antonietti et al. 2015) to these students to assess their abilities and competences. In particular, the SAI test measures: i) basic cognitive functions (such as attention and memory); ii) emotional functions (such as empathy and emotional recognition); iii) self-regulatory functions (such as body awareness). This test was administered in school year 2018/2019, when children attended their second year of pre-primary school (i.e. when they were 4 years old), and in school year 2019/2020, when the same children attended their third year (and they were 5 years old). We then had the opportunity to link these data with administrative data from the Minister of Education and with the results of the INVALSI test administered to students in grade 2 of primary school for the two main subjects (Italian language and Mathematics). Therefore, we obtain a unique database including measures of students' competences in three points in time, two before and one during compulsory education.

Our results clearly show that: i) the cognitive dimension of the SAI test is unequally distributed in our sample, with migratory background being the most relevant factor associated with students' skills in the second year of pre-primary education; ii) the difference between Italian and foreign students virtually disappears at the end of pre-primary school, thus suggesting the role of pre-primary education in reducing inequalities; iii) on average, differences between Italian and foreign students appear once again during primary education, suggesting that primary school may be not as effective as pre-primary school in integrating students with migratory background. Nonetheless, our data show a relevant exception: students with a migratory background and a good level of pre-primary competences tend not only to achieve the competences of Italian students, but also to outpace them – being all other characteristics equal – when reaching grade 2 of primary school.

From a policy perspective, this last result is particularly interesting, considering that, at the national level, the number of foreign children not enrolled in pre-primary education is still very high and that this is one of the factors most strongly associated with differences in students' curricular competences.

Keywords: Educational inequalities, early skills, pre-primary school, migratory background

First-Generation College Graduates (PrimaGen): Inclusion, Support, Empowerment A project of the State University of Milan

Angela Biscaldi – Giuseppe Mazzarino – Barbara Rosina

In recent decades, the experience of first-generation university students—that is, those who are the first in their families to pursue higher education—has attracted growing interest in both academic research and educational policy on an international scale (O'Shea, May, Stone, Delahunty 2024). This segment of the student population, while diverse, shares a common experience: entering the university system without a direct family legacy in higher education. The structural and symbolic implications of this condition give rise to a wide range of challenges, not limited to academic learning, but extending into relational, emotional, economic, and cultural domains.

For first-generation students, the university is not only a place of education; it becomes a transformative space that profoundly reshapes family dynamics, future aspirations, and self-perception within the broader social landscape (Romito 2021).

These experiential complexities were at the heart of the *PrimaGen* project, promoted by the University of Milan as part of the PRO-BEN program (*Universities for Psychological Wellbeing: From Prevention to Intervention*), which is entirely dedicated to student wellbeing. Running from January to September 2025, the project aimed to explore the practices, trajectories, challenges, and resources that shape the academic journeys of first-generation students, offering a nuanced and systemic understanding of their specific experiences.

The research not only addressed overt challenges—such as navigating bureaucratic systems, financial hardships, and social isolation—but also examined the coping strategies students employed, the meanings they attached to their academic paths, and the wider repercussions on their families and communities.

A qualitative, ethnographically grounded methodology was employed. The core of the research consisted of in-depth interviews with eighty first-generation students, supported by contextual analyses of university support systems, social networks, and cultural representations of education. This methodological approach was chosen to foreground the voices and perspectives of those directly involved, avoiding reductive generalizations in favor of a situated and contextualized understanding of experience.

The ethnographic lens enabled the research to capture the relational and transformative aspects of the university journey, showing how it not only alters individual trajectories but also reshapes family relationships, intergenerational expectations, and one's sense of social belonging.

One of the key themes explored by the project was the cultural distance between the university environment and the students' backgrounds. In the absence of family role models, many students faced a form of "cultural translation," requiring them to master not only academic content but also institutional codes, unwritten norms, social interactions, and strategies for navigating university life. This often led to feelings of inadequacy, difficulty in accessing reliable information, uncertainty in decision-making, and a sense of disorientation—factors that could result in delayed progress, changes in course, or even dropping out (O'Shea 2016; Bell, Santamaria 2019).

Yet, alongside these challenges, powerful stories of transformation also emerged. For many, university offered a concrete path to social mobility, a way to overcome difficult family circumstances, and an opportunity to rewrite personal narratives and inspire new collective ambitions. Academic success was not experienced solely as an individual achievement, but often as a shared triumph—a way of giving back to one's family and community. In this sense, the university took on a symbolic role far beyond its traditional function, becoming a space where new forms of belonging were created, evolving identities were affirmed, and students grew not just academically, but also socially, emotionally, and politically (Davis 2010).

The project included interviews not only with students but also with faculty members, tutors, and administrative staff. Of particular significance were the stories of second-generation students who were still the first in their families to attend university. Their narratives shed light on a rarely explored intersection:

high familial expectations, cultural negotiation, and unfamiliarity with the academic environment. These voices allowed for a deeper understanding of the relational and institutional dynamics shaping the academic journey, including how legitimacy in higher education was built—or undermined—how merit was perceived, and which symbolic and practical strategies were adopted to navigate challenges.
Special attention was also given to friendship networks, the use of university services, and the often difficult but generative process of developing a sense of belonging within academic settings.
Another area of research focused on mapping university support services to critically evaluate the effectiveness of existing inclusion policies. This analysis helped identify both best practices and areas in need of improvement, allowing for the development of concrete recommendations to enhance the university's capacity for inclusion. Within this framework, the university was seen not only as an educational institution, but also as a social and political actor responsible for supporting vulnerable pathways, eliminating barriers, and promoting equity.
In this way, *PrimaGen* represented an innovative and multidimensional contribution to understanding the experiences of first-generation students within the Italian higher education system. In a context often marked by educational inequality, the project provided a detailed mapping of the challenges, needs, and specific resources of this student group. Its scientific value lies in filling a gap in the national literature—still underdeveloped on this topic—and in offering a solid empirical basis for both theoretical reflection and practical intervention.
Through its anthropological approach, the project generated deep insights into the dynamics of inclusion, participation, and wellbeing among first-generation students, enriching the academic discourse on equity and social justice in higher education. A distinctive feature of *PrimaGen* was its emphasis on communication, with the creation of visual and multimedia materials for public engagement and awareness-raising.
The use of accessible language and inclusive communication tools allowed for effective dissemination, engaging a wide range of stakeholders—academic, institutional, and civic—in a process of collective reflection.
PrimaGen thus positions itself not only as a project of observation and analysis, but also as a catalyst for transformative visions of a more equitable, welcoming, and socially conscious university, one that values difference as a source of strength

Keywords: First in family/first generation, education, inclusion, mobility

Learning Loss and Recovery in the COVID-19 Timeline: The Role of the Language Spoken at Home

Francesca Carimando – Loris Vergolini – Eleonora Vlach

Introduction

Over the past decades, Italy has experienced a significant demographic shift, becoming a destination for international migration (Ambrosini, 2005; Colombo & Sciortino, 2004). As classrooms become increasingly diverse, addressing educational disparities between native students and children of immigrants has emerged as a policy concern. A persistent ethnic achievement gap has been widely documented in the Italian context, with children of immigrants consistently underperforming compared to their native peers (Contini & Azzolini, 2016). Among the many factors contributing to this disparity, the language spoken at home has been shown to play a relevant role (Triventi et al., 2022).
Building on this, the outbreak of COVID-19 disrupted schooling as classrooms closed and instruction shifted online almost overnight. While the pandemic posed obstacles – reduced interaction with teachers and peers – for all students, evidence shows these challenges deepened existing divides (Bertoletti et al., 2023; Grewenig et al., 2021). A global meta-analysis by Betthäuser et al. (2023) confirms widespread learning losses, especially in mathematics, with the greatest impacts among disadvantaged students. These findings raise urgent question about how quickly and fully all students recovered once schools return to normal.

Research object and hypotheses

Although an emerging body of research has documented the influence of the pandemic on education, most studies focus on differences related to socioeconomic status or gender, with less attention paid to ethnic

disparities. This study addresses that gap by examining how COVID-19 influenced ethnic educational inequalities in Italy, with a particular focus on the language spoken at home.

Drawing on INVALSI data from 2018/2019 to 2022/2023, the analysis compares academic performance in reading and mathematics across four cohorts of students, covering the pre-pandemic (2018/2019), pandemic (2020/2021 and 2021/2022), and post-pandemic (2022/2023) periods. The goal is to assess whether students with a migration background experienced steeper learning losses during the pandemic, how their academic performance evolved once schools fully reopened, and to what extent the language spoken at home influenced both the losses and the recovery trajectory.

It is hypothesised that students with a migration background – especially first generations with an extra-European background – already lagged behind their peers prior to the pandemic, experienced disproportionately larger learning losses under remote instruction, and displayed uneven recovery once in-person teaching resumed. Moreover, the language spoken at home is expected to mediate these patterns, interacting with both migration background and geographic origin, influencing the magnitude of learning losses during remote instruction and once schools reopened.

Data

The study uses data collected by INVALSI on the entire population of students enrolled in grades 5, 8, and 10 during the academic years from 2018/2019 to 2022/2023. These three grades were selected as they represent distinct stages of the Italian school system – primary, lower secondary, and upper secondary education – allowing for an overview of different educational levels. The dataset combines standardised test scores in reading and mathematics with background information derived from student questionnaires. In particular, the language spoken at home is included among the covariates of interest. Since the student questionnaire was administered only in grades 5, 8, and 10 during the selected years, data for grades 2 and 13 were excluded from the analysis.

Method

The analysis focuses on the outcomes of the INVALSI reading and mathematics assessments as dependent variables. Migration background is the primary independent variable and is categories into groups based on the birthplaces of students and their parents. These include: Natives (students born in Italy with both parents born in Italy), Mixed (students with one parent born in Italy and the other abroad, regardless of the student's birthplace), Second-generation (students born in Italy with both parents born abroad), and First-generation (students and both parents born abroad). Within the Mixed, Second-generation, and First-generation groups, further distinctions were made based on the geographical area of origin – namely, European Union (EU), Europe (non-EU), and extra-European countries.

Basic control variables are included, such as gender, socioeconomic status (measured through the ESCS index), year of birth, early education attendance, and the language spoken at home. For the latter, detailed data were available, allowing students to indicate one out of several languages (up to 22). To simplify the home-language variable, the distribution of reported languages was examined, identifying the most commonly spoken: Italian, Albanian, Arabic, Chinese, English, Romanian, and Spanish. The remaining fewer common languages were combined into a single "Other language" category.

The analysis proceeded in two main steps. First, descriptive statistics were examined to provide an overview of student performance in reading and mathematics assessments across the four academic years, disaggregated by migration background. Second, a series of OLS regression models was estimated separately for each grade, academic year, and subject, with migration background as the main independent variable. An interaction term between migration background and language spoken at home was included to explore whether differences in linguistic exposure influenced academic outcomes.

Results and findings

Preliminary findings reveal a similar pre-pandemic pattern across grades 5, 8, and 10 in both reading and mathematics. In 2018/19, native students led performance, mixed-background peers trailed only slightly, and second-generation learners outscored first-generation students – especially those from extra-European origins. Notably, however, in grade 10 mixed non-EU students outperformed their native counterparts in both reading and mathematics. Non-Italian home-language penalties were large: English speakers fared best, while Chinese, Arabic, and Spanish speakers seemed to suffer the deepest deficits.

During 2020/21 and 2021/22, mixed and EU-origin students showed relative resilience, with language gaps widening. First- and extra-European groups, particularly in grades 5 and 8, experienced sharper declines under remote learning. Notably, first-generation Albanian speakers in grade 8 saw their interaction effects worsen in reading, whereas first- and second-generation Chinese speakers exhibited increasingly positive interactions in mathematics.

By 2022/23, recovery was uneven. Older and second-generation students began closing their gaps – most visibly in mathematics – while younger first-generation learners in grades 5 and 8 continued to lag. Interaction effects for Chinese-speaking students remained strongly positive across both migrant generations, whereas Spanish-speakers' disadvantages grew.

Keywords: COVID-19, learning loss, learning recovery, ethnic achievement gap

Double barriers? Gender and migration backgorund in STEM studies access

Cinzia Conti – Eugenia Bellini – Francesca Dota – Massimo Strozza

In Italy, according to the latest data released by the Ministry of University and Research (MUR), women represent nearly 56% of all university enrollments in the 2024/2025 academic year, but less than 38% in STEM fields. Additionally, both national and international literature on migrants' descendants highlight the significant challenges faced by foreign students during the transition to university studies (Buonomo et al., 2024).

A study conducted in Canada by Finnie and Child (2018) examines the factors influencing Canadian students' decisions to pursue STEM (Science, Technology, Engineering, and Mathematics) fields. The authors analyze demographic, socioeconomic, and academic variables that contribute to these career choices. Key findings show that students' participation in STEM is strongly influenced by academic preparation in math and science during high school, parental education levels, and family income, with notable gender- and ethnicity-based differences. This research highlights how early academic and socioeconomic conditions shape students' pathways into STEM disciplines.

Another relevant study by Gutfleish and Kogan (2022) emphasizes that students' STEM achievements are influenced by their parents' socioeconomic status and occupational background. The study reveals significant disparities, showing that students from higher-status occupational backgrounds tend to perform better in STEM, with gender and ethnic origin further affecting outcomes.

Based on these insights, the hypothesis of the present study is that, in Italy, both gender and citizenship play an important role in the decision to pursue a STEM degree, and that foreign girls may face a double disadvantage.

This research is based on a longitudinal dataset created through record linkage of the ISTAT Survey on Second-Generation Students (2015), the Population Register, and administrative university student data (academic years 2015–2016 to 2022–2023). The study aims to offer an initial contribution to identifying individual and family factors influencing the educational choices of second-generation students, with a particular focus on gender and citizenship.

The longitudinal analysis of integrated data shows that gender disparities are clearly visible across all categories. Overall, girls interviewed in secondary schools in 2015 show a higher likelihood of being enrolled in university than boys (58.6% vs 42.0%), a result that is particularly pronounced among Italian students (60.5% for women vs 44.2% for men). However, this female advantage drops sharply in STEM fields, where only 11.2% of women are enrolled compared to 18.6% of men. This suggests persistent gender segregation in academic disciplines, with women underrepresented in STEM despite their overall higher university participation.

Ethnic disparities are also evident. Foreign citizens have significantly lower enrollment rates than Italians, both overall and within STEM and non-STEM fields. For instance, only 34.8% of foreign students are enrolled in university, compared to 52.0% of Italian students. The gap is even more striking among foreign males, whose enrollment rate is only 25.0%. This suggests that foreign students, particularly males, face structural or socioeconomic barriers in accessing higher education.

The combined analysis of gender and ethnic background reveals cumulative disadvantages: foreign girls are doubly marginalized in STEM fields, with an enrollment rate of just 8.5%, the lowest among all groups considered. Conversely, Italian women are the most represented in non-STEM courses (48.9%), indicating a strong gender-based preference or steering toward certain disciplines.

A logistic regression model was used in the analysis to explore the relationship between various individual, family, and contextual factors and students' educational choices, as suggested in the literature. The dependent variable is the choice to enroll in a STEM or a non-STEM course. The model included only

students who were attending upper secondary school at the time of the interview and who enrolled for at least one year in an Italian university between 2015 and 2022.
The results of the logistic regression model provide a clear picture of the factors influencing students' educational choices. Among all the variables considered, gender and citizenship emerge as particularly significant in shaping these trajectories.
Starting with gender, the effect is both substantial and highly significant. Female students, compared to their male peers, are much less likely to reach the outcome considered. The odds ratio is 0.250, indicating that, all else being equal, girls have 75% lower odds than boys. This is a striking result, pointing to a persistent gender gap and suggesting that girls may face specific barriers that limit their progress or aspirations within the education system.
Citizenship is also a key variable. Compared to Italian students (the reference group, which is not representative of the entire student population), all the foreign groups studied show a higher likelihood of enrolling in STEM courses. In particular, students of Chinese origin are the most likely to choose STEM, with an odds ratio of 1.709, suggesting they are over 70% more likely than Italians to do so. It is important to note that these are Italian students from the control group and do not represent the entire population of Italian upper secondary students. Students of Albanian and Romanian origin also show significantly higher odds, indicating a clear tendency among certain foreign groups to pursue STEM studies. Other important factors include school background and family characteristics.
In sum, this work highlights the complex interplay between gender, migration background, and academic preparation in shaping access to STEM education. Addressing these disparities requires targeted policy measures, more effective guidance in secondary schools, and broader efforts to support the educational aspirations of second-generation students, particularly girls.

Keywords: Gender, migration background, university, longitudinal approach

Bridging the Gap? Educational Achievement Trajectories of Migrant Generations in Italy

Sara Fiasconaro – Moris Triventi

Educational inequalities affecting students with a migration background remain a pressing issue in contemporary societies and a key challenge to achieving Sustainable Development Goal 4 on inclusive and equitable education. While cross-sectional studies have extensively documented the so-called "migration-related achievement gap," they often adopt a static and binary conceptualization of migrant status, neglecting both the temporal dynamics of educational trajectories and the internal heterogeneity of migrant groups.
This study draws on several theoretical perspectives to frame its analysis and develop hypotheses. Classic assimilation theories suggest a gradual convergence of educational outcomes across migrant generations, driven by increased language fluency, cultural integration, and familiarity with the school system. However, more recent theories, such as segmented assimilation and social stratification approaches, highlight the role of structural and institutional barriers that hinder the full integration of some migrant groups. According to these perspectives, factors like socioeconomic disadvantage, ethnic discrimination, and differential treatment by schools may contribute to persistent inequalities. Furthermore, Boudon's distinction between primary and secondary effects of social origin provides a valuable lens to interpret how initial academic performance and educational choices interact in shaping long-term outcomes. The concept of compensatory advantage also informs the hypothesis that highly educated families may buffer the negative effects of migrant background. These theoretical elements jointly support the expectation that achievement gaps may persist over time and vary across groups and contexts.
This study provides new empirical evidence on the evolution of academic achievement gaps across different migrant generational groups in Italy, using a longitudinal approach. Specifically, it focuses on how students' relative academic positioning in mathematics and language proficiency changes across the educational career—from the end of primary school to the final year of upper secondary education. The empirical analysis compares native students with three categories of migrant-origin students: the 1.5 generation

(foreign-born, arrived in Italy after age 6), the 1.75 generation (foreign-born, arrived between 0 and 5 years), and second-generation students (born in Italy to two foreign-born parents).
The thesis uses high-quality administrative and survey data from the INVALSI National Student Assessment System (SNV). Two cohorts of students are observed longitudinally at four key stages: Grade 5 (end of primary), Grade 8 (end of lower secondary), Grade 10 (beginning of upper secondary), and Grade 13 (end of upper secondary). The construction of the longitudinal dataset involved careful harmonization procedures to ensure consistent identification of students' generational status, social background (parental education), and other control variables such as gender.
Two main analytical strategies are adopted. First, growth curve models are employed to examine the development of standardized test scores over time, estimating both initial gaps and their evolution throughout the school career. These models allow for assessing whether achievement trajectories among different generational groups converge, diverge, or remain stable over time. Second, the Karlson-Holm-Breen (KHB) method is applied to decompose the total effect of migrant background on Grade 13 achievement into direct and indirect components. Specifically, the analysis examines whether the observed gaps in final performance are mediated by early achievement (Grade 5 scores) and by track choice at the end of lower secondary education (i.e., academic vs. vocational pathway).
Results indicate persistent and substantial achievement gaps between migrant-origin students and their native peers. Students from the 1.5 generation show the most disadvantaged trajectories, particularly in language, and remain significantly behind natives at all stages. Second-generation and 1.75-generation students perform better but still fall short of parity with native students. There is only limited evidence of convergence over time, suggesting that initial disadvantages are not automatically compensated during the educational career.

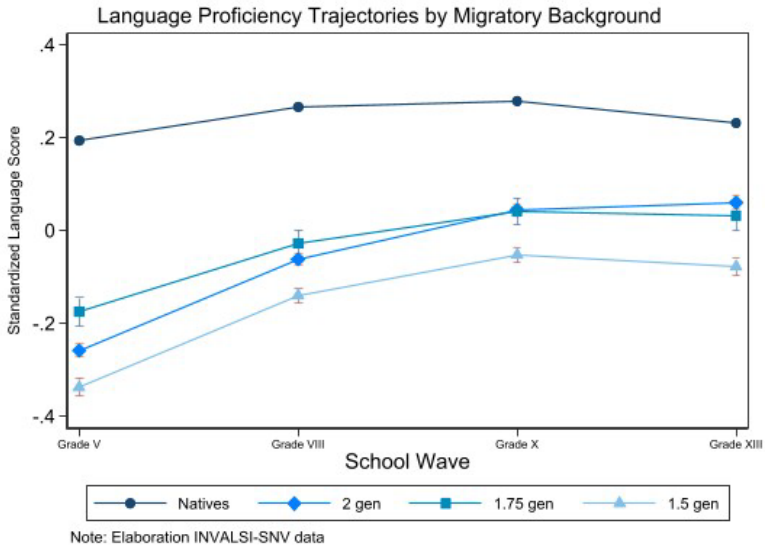


Fig 1 Evolution of Relative Performance in language by Migratory Background

The growth curve analysis reveals that achievement gaps tend to stabilize rather than close, particularly in mathematics. For language proficiency, some signs of narrowing emerge among the 1.75 generation, but even in this group, full catch-up is not achieved. These patterns challenge the assumptions of linear assimilation and confirm the importance of early educational disparities.
The analysis also uncovers important moderating factors. Parental education plays a significant role: while students with highly educated parents tend to perform better overall, migrant-native gaps are smaller among this group—supporting the compensatory advantage hypothesis. Nonetheless, migrant-origin students with tertiary-educated parents still underperform compared to their native counterparts, suggesting that family resources alone are insufficient to equalize outcomes. Gender also interacts with migrant background: girls outperform boys in language but lag in math, and these gender gaps are more pronounced among natives than among migrant-origin students.
The KHB decomposition highlights the importance of cumulative mechanisms. A substantial share of the Grade 13 migrant-native achievement gap is mediated by earlier test scores and by track choice. In particular, the educational track attended in upper secondary school—strongly influenced by Grade 8

achievement and teachers' recommendations—emerges as a critical driver of long-term disparities. This finding underscores the role of institutional structures, such as early tracking, in perpetuating disadvantage. These results carry relevant implications for both academic research and educational policy. From a scholarly perspective, the study contributes to the literature on educational stratification by combining a life-course and generational approach, highlighting how migration-related inequalities evolve across time and intersect with other dimensions of social stratification. The findings emphasize the need for more nuanced categorizations of migrant background, avoiding the oversimplification of treating migrant students as a homogeneous group.
From a policy standpoint, the study points to the need for early interventions aimed at preventing the accumulation of disadvantage among migrant-origin students. Specific attention should be paid to the period of transition between lower and upper secondary education, where track choices can reinforce or mitigate pre-existing gaps. Policies that support language acquisition, individualized learning support, and greater guidance for families unfamiliar with the school system may be particularly beneficial.
In conclusion, while some signs of progress can be observed, the Italian education system still faces major challenges in providing equitable opportunities to all students, regardless of generational background. Addressing these challenges requires both structural reforms and targeted strategies to support inclusion, particularly for students arriving later in the school system.

Keywords: Educational inequality, Migration background, Longitudinal analysis, School tracking

Agenda 2030: Goals 4 and 10 for a quality and inclusive school starting with invalsi items

Ivan Graziani – Stefano Babini

The 2030 Agenda is a policy document signed in 2015 by the governments of the 193 countries belonging to the United Nations. The document contains a set of goals to be achieved by 2030 to improve the living conditions of our planet and its inhabitants, with special consideration for reducing the wide gaps between peoples.
Education, in particular, is a key element in improving the quality of life and helping a community develop. Although schooling has generally increased around the world, still too many children do not attend school or drop out early, so that they are unable to read, count and write (OECD, 2018).
It may seem easy to imagine, a good level of education in our school in Italy. This is a very important element in being able to improve the quality of students' lives and foster the development of a school community.
Self-Assessment Reports (RAV) with Improvement Plans (PDM) and Three-Year Plans of Educational Offerings (PTOF) of schools have been developed by institutions with commitment and vision for their futures. The themes of sustainable development, Agenda 2030 and global citizenship had already been reinforced in February/March 2018 thanks to the working document "National Directions and New Scenarios," developed by the National Scientific Committee for the accompaniment of the National Directions for preschool and first cycle (DM 254/2012).
Citizenship and sustainability education, the themes of Agenda 2030, and the digital transformation are precisely the new scenarios on the basis of which the Commission rereads and renews the National Directions.
Goal 4 aims to provide all students with quality, inclusive and equitable education and to promote equal learning opportunities for all.
This Goal addresses all students young and old. Indeed, it is very important that young people learn all the knowledge and develop all the skills that will enable them to become adult citizens capable of making choices and building their own futures, choosing the jobs they want and improving their living conditions.
Goal 10 examines the inequalities present in the world, but also in our schools. Indeed, it is crucial to promote inclusive growth starting right from schools, regardless of age, gender, ethnicity, religion and anything else that may hinder proper learning (UN, 2015).
In our schools there are several foreign students or students with other difficulties with various schooling problems.
It is precisely this aspect that we wanted to address in our research.

The method used was mixed, and our intervention was developed at the beginning of the year, October 2024, in the second secondary classes of grade I and the first secondary classes of grade II.
The first goal we set ourselves was to offer themed lessons with a holistic, inclusive approach to engage multiple styles of intelligence and learning (Falcinelli, Gaggioli, Capponi, 2016).
The different themes were identified and developed in a logic of continuity that took into account the growth of learners and their learning needs.
The second objective was to work on skills, fundamentals with a focus on functional literacy, math literacy and multilingual competence. Each student got to know their skills and how to advance to the next level. The focus on student-centeredness with individualized and personalized actions enabled all students to maximize educational success.
The third goal was to leverage students' experiences as they were involved in real-world hands-on projects, deepening their understanding of their place in the mathematical world.
The main goal was to improve their ability to solve unfamiliar problem situations by observing the development of students who had struggled in the entry test.
Our aim was to be able to ensure that all students acquired the knowledge and skills necessary to promote sustainable development, aimed at respect for human rights, gender equality, promoting consolidated learning and skills, and valuing cultural diversity (De Carlo & Pugacheva 2021).
To do this, we selected items on Gestinv3.0, from among those that had responses above 50 percent, and then administered a fascicle with those items that did not contain questions that had not yet been addressed, and then in April 2025, in the same classes, In April 2025, we administered different items, but with similar characteristics, to see if the teaching had a positive impact to reduce the gaps found in the October 2024 administration in the two institutions.
We were able to observe, in the test administered in April 2025, a positive evolution of results compared to the October administration, as comprehension of the text, as the level of skills gained. In particular, the results were 24 percent better in grade I secondary schools and 28.5 in grade II secondary schools.
Most students improved their test results, and students with various frailties improved in both I and II grade.
A very positive fact was the fact that all students participated, regardless of income, age, gender, religion, certified difficulty or geographical origin. This was a good start to be able to ensure that everyone has the same opportunities for study and learning.
Students did, in fact, confront each other, spontaneously at the end of the two tests, with reasoning about the difficulties they encountered and the parts that were less clear. We then engaged in a whole-class discussion in the classes to be able to talk about math and the difficulties.
Students then asked to use self-assessment tools, already used in other subjects, such as rubrics and checklists. They built them, with our guidance to foster authentic assessment and the role of technology in fostering learning.
In conclusion, in the facts, sustainability is often understood as an aspect of growth, but it cannot be understood only in identifying new profit classes, but we must learn to act on problems related to information, education and lifelong learning, to be started early in the two secondary schools (Iavarone et al, 2017),

Keywords: Goal 4, teaching action, invalsi items, feedback

Bridging or Widening the Gap? The Role of School-Level Resources for Migrant and Native Student Achievement in Italy

Elisa Sbalchiero – Emanuele Fedeli – Moris Triventi

Competencies acquired throughout the academic journey can significantly influence the subsequent school-to-work transition, thereby shaping early career trajectories and affecting future occupational attainments. The importance of academic achievement for broader social and labour market-related outcomes is particularly evident when examining the life trajectories of immigrant-origin students, given the implications for their socio-economic integration into the host society (Muñoz-Comet & Arcarons, 2022; Heath & Brinbaum, 2014). Recent research on academic careers has started to explore how school-level factor or schooling contexts contribute to unequal achievement outcomes among students, including within the Italian education system. In this study, we will examine the role of schools in look in the development of achievement gaps, focusing specifically on disparities in academic performance between students with and without a migration background during early upper secondary education in Italy. Italy is considered a relatively recent country of immigration compared to other European national contexts. The proportion of second-generation immigrants or migrant descendants has risen noticeably in recent years, particularly within upper secondary schools (ISMU, 2018; Ceravolo & Molina, 2013; Azzolini, Schnell, & Palmer, 2012).
School-related factors, such as the socio-demographic composition of the student population, may influence educational and learning outcomes (Minello & Barban, 2012). The proportion of migrant-origin students or those from diverse social backgrounds has been linked, even negatively, to the development of educational aspirations and levels of academic engagement (Kuyvenhoven & Boterman, 2021; Fekjær & Birkelund, 2007). In explaining differences in academic achievement, another critical school-level factor is the availability or quality of material and human resources. These resources shape the learning environment in the schools and define the type of *stimuli* to which students are exposed, potentially affecting their academic outcomes (Cebolla-Boado & Medina, 2011). This study aims to contribute to the existing literature by exploring the role of school-level resources in explaining academic achievement disparities between migrant-origin and native students. Specifically, it seek to answer to the following research question: *to what extent do school-level educational resources play a role in mitigating the academic achievement gap between students with and without a migration background?*
Literature on the role of schools in shaping educational inequalities highlights the beneficial effects of the schooling environment on students' learning outcomes, including those with a migratory background (Downey & Condron, 2016; Raudenbush & Eschmann, 2015). Studies emphasized the compensatory and equalising functions associated to schools, framing their role as potentially reducing differences in academic achievement along students' educational trajectories (Sprong & Skopek, 2023). Academic achievement is also characterised by a path-dependent nature: initial advantages or disadvantages in performance may be reproduced and even strengthened over time (Passaretta & Skopek, 2018). However, students from disadvantaged backgrounds may still benefit from school-based learning opportunities. As they progress across educational levels, their achievement levels may improve and existing disparities in academic performance can be reduced or gradually equalised (Sprong & Skopek, 2023; Passaretta & Skopek, 2018). Accordingly, increasing the availability of educational resources within schools is associated with higher levels of academic achievement. These school-level factors may contribute to reducing disparities in academic competencies between migrant-origin and native students (Downey & Condron, 2016; Raudenbush & Eschmann, 2015).
In this study, we account for the varying levels of resources across schools and for their role in shaping the academic achievement disparity among migrant-origin and native students by employing a multilevel method of analysis. We consider both material and immaterial educational resources within the school context and these refer to the infrastructure conditions, availability of digital devices, teaching and staff quality. We make use of a unique longitudinal dataset that merges PISA (an age-based survey which provides information on individual characteristics, school context, and achievement for 15-year-old students) data from the School Questionnaire with student academic performance scores from INVALSI (Grade 10). We were able to overcome a key limitation in previous studies by including and linking to students in early upper secondary education a measure of their prior performance from INVALSI (Grade 8). We are, thus, able to provide a more accurate estimation of how academic achievement disparities related to migration background evolve over students' educational trajectories, particularly in the transition to upper secondary education.

Preliminary results indicate that students with a migration background seem to show a relative advantage in reading competencies compared to natives within schooling environments having scarce resources. Whereas, in well-equipped schools, migrant-origin students may not benefit from higher extent of educational resources by improving their academic performance, as it happens for their native peers, exhibiting lower reading competencies in comparison to them. These findings can provide insights into how, and to what extent, school contexts may contribute to either mitigating or further exacerbating academic achievement gaps by migration background. In doing so, they could potentially contribute to the literature on migration-related educational inequalities and the broader integration trajectories of students with a migratory background within the host country's society and economy.

Keywords: Educational achievement, students with a migration background, school resources, multilevel method

SESSION 5. ASSESSING DIGITAL SKILLS AND ASSESSING LEARNING WITH DIGITAL: NEW CHALLENGES FOR SCHOOLS

ORGANIZER: INVALSI - UNIVERSITY OF BERGAMO
COORDINATOR: PAOLO BARABANTI - MARCO GIGANTI
20TH NOVEMBER: 8.30 A.M. – 10.45 A.M. {AULA 4 – RESEARCH 8}

Insights from the IEA Computer and Information Literacy Study, 2023

Julian Fraillon

Introduction

The International Association for the Evaluation of Educational Achievement (IEA) has been studying the use of ICT in school education since the late-1980s (Pelgrum & Plomp, 2011). IEA initiated the International Computer and Information Literacy Study (ICILS) in response to increasing recognition, from late last century, of the necessity for young people to develop digital literacy competencies to support them to function effectively as learners, workers and participatory citizens in an increasingly digitalized world (see, for example, Pedersen et al.,2006, European Commission, 2008, Ferrari, 2012). The first cycle of ICILS was conducted in 2013 with subsequent cycles being conducted in 2018 and 2023. The next cycle of ICILS data collection will take place in 2028.

ICILS assesses how well grade 8 students are prepared for life in a digital age. Specifically, it examines students' abilities to use ICT productively across a range of purposes, emphasizing critical thinking, evaluation, communication and problem-solving skills, that extend beyond basic digital functions. ICILS 2023 was built on previous cycles, extending the time series and scope of investigation to monitor global trends in digital literacy.

ICILS 2023 included a test of students' computer and information literacy (CIL), defined as: an "individual's ability to use computers to investigate, create, and communicate in order to participate effectively at home, at school, in the workplace, and in society" (Fraillon & Duckworth, 2024, p. 26).

ICILS 2023 also included, an optional test of computational thinking (CT), defined as "individual's ability to recognize aspects of real-world problems that are appropriate for computational formulation and to evaluate and develop algorithmic solutions to those problems so that the solutions could be operationalized with a computer" (Duckworth & Fraillon, 2024, p. 38).

ICILS 2023 further investigated the characteristics of students, their access to and use of digital technologies, and the contexts in which students' CIL and CT develop. These contextual data were collected from national study centers, schools, teachers and students.

ICILS includes two achievement scales, one for CIL and one for CT. These scales were established in 2013 and 2023 respectively and. The scales provide measures and concomitant descriptions of the nature of student achievement across different levels of proficiency. The scales provide the empirical bases for analysing and reporting student achievement. This includes analyses of differences in achievement among countries and among other subgroups of students, and of the relationships between contexts in which CIL and CT develop and CIL and CT achievement.

In this session, I will present key findings from ICILS 2023 with respect to students CIL and CT achievement, their experiences with ICT in and outside the classrooms and the broader implications of these findings for life in the digital age.

Data

ICILS 2023 collected data from 132,998 grade 8 (or equivalent) students in 5,299 schools across 34 countries and one benchmarking participant. These student data were augmented by data from 60,835 teachers in those schools, and by contextual data collected from school ICT coordinators, principals, and national research centers. Twenty-four countries, and one benchmarking participant also participated in the optional CT assessment.

Method

ICILS 2023 used a complex sampling design including "multi-stage sampling, stratification, and cluster sampling" (Tieck, 2025, p.71). From within each country (or benchmarking entity) schools were selected for participation using probability proportional to size (PPS) sampling with stratification. Within each school, typically one class of Grade 8 students was sampled to participate. Up to 15 teachers of students in

Grade 8 were also sampled within each participating school. The ICT-coordinator and principal within each sampled school were also surveyed. Full details of the sampling procedures are reported in the ICILS 2023 Technical Report (Tieck, 2025).

Each sampled student completed ICILS on computer (either using the internet or USB-based delivery). Students completed a one-hour test of CIL, comprising two 30-minute test modules randomly allocated to students from a set of seven modules, in a fully balanced rotational design. Each module "comprised a set of questions and tasks based on a real-world theme" (Fraillon, 2024, p. 3). Each student then completed the student questionnaire designed to be completed in approximately 25 minutes. In countries also completing the CT option, students then had a short break before completing the 50-minute CT test. The CT test comprised two 25-minute test modules randomly allocated to students from a set of seven modules in a fully balanced rotational design. The CT modules comprised "questions and tasks relating to real-world problems that may be addressed with computer-based solutions" (Fraillon, 2024, p. 3).

Sampled teachers completed a 30-minute online questionnaire, and the ICT-coordinator and school principal each completed 20-minute online questionnaires.

This proposal reports on ICILS 2023 data collected from students and teachers.

The ICILS 2023 data were cleaned and processed by the ICILS international study center (ISC). The CIL and CT scales were established using Rasch IRT modelling and equated to the existing ICILS CIL and CT scales. Questionnaire response data were reported as proportions of response by category and, where planned and supported by the data, Rasch IRT scales were created using groups of questionnaire items. In all cases, rigorous evaluation of the measurement properties, including the cross-national applicability (through evaluation of measurement invariance) of all scales used in ICILS 2023. Full details of the test and questionnaire scaling methodology are provided in the ICILS 2023 Technical Repot (Liaw, Rožman, Fraillon & Gonzalez, 2025; Chen, Christiansen, Rožman, & Rolf Strietholt, 2025).

A key aim of ICILS is to report unbiased measures of population estimates within each participating country. As is now standard practice in ILSA, to achieve this aim, the ICILS 2023 ISC applied weights and corresponding estimates of the uncertainty introduced by the ICILS complex sampling design. Plausible value methodology was also applied to provide estimates of the uncertainty introduced by the ICILS incomplete block design in which each student completed a sample of the total available CIL and CT test content. Full details of these processes are provided in the ICILS 2023 Technical Report (Strello, Rožman & Fraillon, 2025).

Results and Findings

This proposal focuses on student CIL achievement from ICILS 2023 together with selected findings associated with students' use of computers at school and outside of school, their confidence to use computers for a range of purposes and teachers' feelings of preparedness to teach aspects of CIL. Following is brief a summary of the results that will be reported in detail in the session.

Student CIL achievement in ICILS is reported across five levels of proficiency from "Below level 1" through to "Level 4". Students working below Level 2 can complete only simple routine tasks on computer and require support when completing information gathering and production tasks. On average across countries, nearly half of students' CIL achievement was below Level 2. The proportion of students below CIL Level 2 ranged from 96 percent (Azerbaijan) to 27 percent (Korea, Rep. of).

The majority of students' capacity to evaluate and verify the accuracy of digital information is very low. When presented with a webpage advertisement that contained four clear markers that the information on the website may be unreliable, on average across countries, 16 percent of students could identify one of the four indicators of unreliability. This ranged from 1 percent (Azerbaijan) to 50 percent (Korea, Rep. of).

Similarly, when asked to state two actions that could be completed to verify the accuracy of information on a website, on average across countries, 24 percent of students could identify one action, and 4 percent of students could identify two.

In contrast, however, students are confident in their capacity to complete such tasks. On average, across countries, 84 percent of students reported that they could moderately well or very well judge whether or not they can trust information they find on the internet.

Furthermore, on average across countries, higher proportions of students reported that they have learned how to complete a range of CIL-related tasks outside of school than at school. Of particular note, with respect to evaluating the reliability of information on the internet. On average across countries, 68 percent of students reported that they have learned to a moderate or large extent how to do this outside of school in comparison to 63 percent of students at school.

Teachers too recognise that this is an area of challenge when dealing with students. On average, across countries, nearly two-thirds of teachers reported believing they needed more professional learning to support students' capabilities to evaluate the reliability of internet-based information.
The ICILS 2023 database is an enormously rich repository of data with respect to aspects of students' digital literacy achievement and the contexts in which students are learning the essential skills for participation in the digital age.
The data summarized in this proposal paint a somewhat bleak picture of the reality of students' CIL achievement in 2023. Very high proportions of students are unable to work independently using ICT to effectively search for, evaluate and communicate with digital information. These competencies are of significance across Europe where the European Commission Resolution on a strategic framework for European cooperation in education and training towards the European Education Area and beyond (2021–2030) (European Commission, 2021), includes a target that less than 15 percent of grade 8 students will have CIL (as measured by ICILS) below Level 2 by 2023. As measured in ICILS 2023, the average across participating European countries was 43 percent.
Of further concern are the data showing a clear mismatch between students' self-beliefs about their capacity to evaluation digital information, and the reality of their objectively measured achievement. Furthermore, students report believing that they are developing some of these skills to a large extend outside of school than within school. This raises questions about what students believe they are learning and how one can better align their uses of technology with more effective critical evaluation of digital information. Teachers clearly recognize this challenge as they report needing more professional learning support in this area.
The ultimate intention of this proposal is to shed light on some areas of specific need that are revealed in the ICILS 2023 data, and to provide researchers and policymakers alike with impetus to make further use of ICILS data to identify and seek solutions to the challenges we face in preparing young people for their lives in an increasingly digital world.

Keywords: Digital competence, Digital literacy, ICT literacy, International large-scale assessment

Digital vs. Traditional Classrooms: An Open Match Between Data, Didactics, and Inequalities

Brunella Fiore

In recent decades, the Italian education system has undergone a profound transformation, driven by the need to adapt to an increasingly digital and interconnected society (Avvisati et al., 2013; Rettore & Checchi, 2014). In this context, the integration of information and communication technologies (ICT) into school settings has taken on a central role in educational innovation processes. With the introduction of the National Digital School Plan (PNSD), established by Law 107 of 2015, the Ministry of Education initiated a process aimed at rethinking the organization of teaching from a digital perspective, promoting more flexible, dynamic, and inclusive learning environments. However, despite substantial infrastructural and training investments, the scientific debate remains lively regarding the actual effectiveness of digital classrooms compared to traditional methods, particularly in terms of their impact on students' academic achievement (Argentin & Gerosa, 2016; Falcinelli & Gaggioli, 2017; Abbiati et al., 2022).
This study fits into the broader body of research aimed at measuring the concrete effects of technological innovation on the quality of learning (Bingimlas, 2009; Bester & Brand, 2013; Delgado et al., 2015; Chauhan, 2017). The analysis focuses on a comparison between digital classrooms, where the daily use of personal devices (primarily tablets) is fully integrated into teaching, and traditional classrooms, where instruction follows conventional methodologies using paper-based materials (Culén & Gasparini, 2012; Cumming, Strnadová & Singh, 2014). The research adopts a longitudinal perspective, analyzing two student cohorts at different stages of their educational path. The first phase, referring to the 2021/22 school year, involved approximately 190 students in the third year of lower secondary school at a comprehensive institute in Lombardy. The 9 classes observed were divided into 3 digital and 6 traditional classes. The INVALSI test scores from Grade VIII were matched with the same students' results from Grade V of primary school. The second phase of the research, currently underway, involves the observation of a new cohort in 2024/25 at the end of upper secondary school, thus enabling an extended longitudinal analysis across four key school transitions.

Since the early 2000s, the adoption of digital technologies in education has generated a growing body of international and national literature. Several studies (Balanskat et al., 2006; Balanskat, 2009; Kong, 2014; Archer et al., 2014; McNaughton et al., 2022) have demonstrated how digital tools can positively influence motivational aspects, student engagement, and perceived self-efficacy. However, other works have pointed to a lack of significant impact on academic performance or, in some cases, even negative outcomes (Sheppard, 2011; Faber et al., 2017). The impact of digital classrooms can vary considerably depending on the subject: while traditional classes tend to perform better in Mathematics, digital classes show positive trends in English—especially in Listening—though these are not always statistically significant (IRVAPP, 2012; Zierer, 2019; Fiore, 2024).

One of the critical aspects highlighted in the literature concerns the role of students' socio-economic and cultural background. The ESCS index (Economic, Social, and Cultural Status), measured through variables such as parents' education level and the availability of educational resources at home, remains a strong predictor of learning outcomes, especially in language and mathematical subjects (Dundar & Akcayir, 2012; Decataldo & Fiore, 2018). However, the impact of ESCS appears to diminish in English Listening assessments, suggesting that access to authentic multimedia content—often integrated into digital learning pathways—may partially compensate for initial disadvantages (Eng, 2005).

From a methodological standpoint, this study uses linear regression models to isolate the effect of classroom type (digital vs. traditional) on INVALSI test results. Control variables include gender, ESCS, the language spoken at home, and previous standardized test scores. The data analysis shows that in Mathematics, traditional classes achieve significantly higher results, even after controlling for initial conditions. In Italian, no significant differences emerge, whereas in English Listening, digital classes exhibit a slight advantage. This trend could be linked to the increased opportunities for repeated, personalized, and interactive listening provided by digital content (Wang et al., 2022). The data also suggest that students with initially weaker performance tend to benefit more from the digital approach in areas such as oral English, where technology may support more gradual and less stressful learning experiences.

Another point of interest is the strong digital orientation of the observed school, which has implemented several innovative projects—such as coding, educational robotics, and flipped classrooms—in alignment with the recommendations of the "Avanguardie Educative" network. However, the mere presence of digital tools does not automatically translate into innovative teaching practices. In some cases, the risk is that of a simple transposition of printed textbooks to digital screens, without a real methodological shift. This raises important questions about teacher training and school leadership's role in supporting a genuine transformation of the learning environment.

A further limitation concerns the difficulty of monitoring the actual use of technology in the classroom: how much time is genuinely devoted to using tablets? How are they integrated into the curriculum? How does the teacher-student interaction change? Without answers to these questions, it is difficult to definitively attribute observed outcomes to the effectiveness of digital learning itself. Moreover, the analyses do not account for variables such as transversal or socio-emotional skills, which are increasingly central in contemporary educational paradigms but are difficult to capture with standardized tools like the INVALSI tests (Liu, Lin & Paas, 2014).

Finally, a relevant sociological issue emerges: the choice made by families to enroll their children in digital or traditional classes is not neutral. Data show a slight prevalence of students with higher ESCS in traditional classes, suggesting that families with greater cultural capital may prefer educational formats that reflect their own experiences. This factor risks introducing a selective bias against digital classes, which are often perceived as experimental or less established.

In conclusion, this study makes a significant contribution to the debate on educational innovation in Italy by offering a rigorous empirical analysis based on longitudinal data. The results underscore the need to assess the effectiveness of technology in relation to subject-specific factors, student profiles, and the actual pedagogical practices adopted in the classroom. Digital integration should not be viewed as an end in itself but as part of a complex educational renewal process that requires continuous training, school leadership, and a shared systemic vision.

The observation of the second cohort, scheduled for 2024/25, will allow researchers to verify whether the observed trends are confirmed over time and whether prolonged use of technology leads to more stable and significant outcomes. In particular, it will be interesting to assess whether the gaps in English Listening widen or narrow, whether digital classrooms manage to close initial performance gaps in Mathematics and Italian, and whether differences emerge in students' subsequent academic and professional pathways.

Overall, the collected data offer valuable insights for policymakers, school leaders, and educators, who are called upon to design more effective and inclusive learning environments. In an era when digitalization is advancing across all sectors of society, it is essential for schools not to passively adopt technological innovations, but to interpret and manage them with pedagogical competence and critical awareness.

Keywords: Digital classrooms, INVALSI standardized test, didactic innovation, educational inequalities

Detecting Differential Item Functioning in INVALSI's 2025 Pilot Assessment of Digital Competences

Luca Bungaro – Chiara Rita Cunsolo – Tommaso Giorgi – Patrizia Falzetti

In 2025, the Italian National Institute for the Evaluation of the Education and Training System (INVALSI) initiated a pioneering pilot program for the assessment of digital competences among Grade 10 students. This initiative represents a major innovation in the national evaluation landscape, aiming to align assessment practices with the European DIGCOMP 2.2 framework and to establish a data-driven foundation for future policy and curricular development in digital literacy. The pilot was administered to a nationally representative sample of approximately 20,000 students and constitutes the first large-scale attempt in Italy to systematically assess digital skills in school-aged populations.

The assessment targeted four out of the five DIGCOMP 2.2 competence areas, Information and Data Literacy, Communication and Collaboration, Digital Content Creation, and Safety, excluding Problem Solving in this initial implementation to streamline item development and ensure construct clarity. A multidisciplinary team developed a pool of 288 original items, classified into three levels of difficulty (basic, intermediate, advanced) and two cognitive categories (knowledge and skill). These items were evenly distributed across the four competence areas and designed to reflect the range and structure of the DIGCOMP 2.2 framework. Test forms were generated using Automatic Test Assembly (ATA) procedures, resulting in 22 distinct forms composed of 30 items each. The ATA ensured internal balance and psychometric equivalence across forms with respect to difficulty level, competence area, cognitive demand, and item interdependence (considering friendly/enemy item dynamics). All forms were delivered digitally, and student responses are being prepared for psychometric modeling and evaluation.

Following data collection, item calibration is expected to be performed using the Rasch model (Rasch, 1960), a unidimensional item response theory (IRT) model that provides a robust framework for estimating item difficulty and student ability on a shared scale. The calibration process is to be conducted using ACER ConQuest software (Adams, Wu, & Wilson, 2015), enabling subsequent analyses to leverage these estimations to explore item performance across demographic subgroups.

The focus of this research project centers on the detection and analysis of Differential Item Functioning (DIF), a key step in evaluating the fairness of standardized assessments. DIF occurs when individuals from different groups, matched on the same underlying ability, have different probabilities of responding correctly to a given item. Identifying DIF is essential to ensure that test items function equivalently across diverse subpopulations and that no group is unfairly advantaged or penalized by the test design.

The analysis will investigate potential DIF across multiple dimensions of student background, including gender, geographic location, socioeconomic background, or type of origin (Italian vs. non-Italian). Given the multidimensional nature of these variables, DIF detection will follow a tiered strategy. For each focal variable, we aim to identify whether items exhibit uniform DIF (i.e., systematic advantage or disadvantage across all ability levels) or non-uniform DIF (i.e., varying item bias depending on student ability). Although the specific techniques for DIF detection will be chosen in the analytic phase, common and well-established approaches include the Mantel-Haenszel procedure, logistic regression modeling, and item characteristic curve (ICC) comparisons under IRT models (Zumbo, 1999; Camilli & Shepard, 1994).

It is expected that DIF analyses will be conducted on items retained after Rasch calibration and psychometric cleaning, which includes evaluation of infit statistics and item-rest correlations. Only items demonstrating adequate fit (e.g., infit between 0.8 and 1.2) and internal consistency (item-rest correlation ≥ 0.2) will be considered, ensuring that DIF detection is performed on a stable measurement base.

The research aims to address the following core questions:

- Do any items function differently across student subgroups after controlling for ability?
- Are there patterns of DIF that suggest systematic bias in specific competence areas?
- How might the presence of DIF affect the validity and equity of interpretations drawn from the assessment?

While the study is currently in its pre-analytic stage, we anticipate that some items may exhibit DIF, particularly in areas where digital access, exposure, or prior experience may differ across social or demographic lines.

Addressing DIF is more than a technical necessity, it is central to the ethical and policy relevance of educational measurement. If unaddressed, DIF can distort our understanding of student competence and reinforce existing educational inequalities. Conversely, identifying and correcting DIF enhances the interpretability and fairness of test results and supports the design of more inclusive assessment instruments.

Beyond the immediate psychometric implications, this component of the project represents a broader commitment to equity in educational assessment. By explicitly examining whether and how item functioning varies across subgroups, the project integrates fairness considerations into the core of test validation. The outcome of this analysis will inform future item development practices, particularly in relation to culturally responsive and accessibility-aware design principles.

In addition, the DIF analysis may serve as an empirical basis for future refinement of the DIGCOMP 2.2-aligned framework in the Italian context. As national and international frameworks evolve, maintaining alignment with the lived experiences and competencies of all students will be essential to sustaining the relevance and credibility of digital competence assessments.

This work thus positions itself at the intersection of psychometrics, educational equity, and digital pedagogy. It advocates for a paradigm in which validity and fairness are inseparable goals, especially in assessments that aim to capture 21st-century skills in a heterogeneous student population. The expected contribution of this study is twofold: providing actionable insights into item bias and laying a methodological foundation for more equitable digital assessments in future.

In conclusion, this study of the INVALSI 2025 pilot assessment project underscores the importance of fairness in large-scale digital testing. By interrogating item-level functioning across critical student subgroups, it not only advances the technical quality of the assessment but also reinforces the social responsibility of public evaluation systems. As Italy takes its first step toward institutionalizing digital competence assessment, analyses such as this are crucial in ensuring that these efforts are inclusive, valid, and capable of supporting all learners equally in a rapidly digitizing world.

Keywords: IRT, Digital Skills, Education, Differential Item Functioning

Academic Achievement of Grade 8 Students and Digital Competence: Evidence from a Joint Analysis of INVALSI and ICILS Data

Luca Bungaro – Marta Desimoni – Mariagiulia Matteucci – Stefania Mignani

This study investigates the relationship between digital competencies and academic performance among lower secondary students (grade 8) in Italy by integrating data from two large-scale international and national assessments: the International Computer and Information Literacy Study (ICILS) and the Italian National Assessment system managed by INVALSI. Specifically, we analyze whether students' digital skills, measured through the Computer and Information Literacy (CIL) scale in ICILS 2018 and 2023, are associated with their achievements in Italian, Mathematics, and English (Reading and Listening), as assessed by the INVALSI standardized tests in the same years.

The core purpose of this research is to explore the extent to which digital proficiency correlates with performance in conventional school subjects, and to what degree this relationship is influenced by student- and school-level factors. Our focus on grade 8 students allows for early insight into how digital literacy intersects with general education outcomes, potentially informing both pedagogical approaches and educational policy.

To conduct this investigation, we perform a precise merging of the ICILS and INVALSI datasets using unique identification codes available in both sources. This data integration enables a rare opportunity to jointly analyze information from two comprehensive and methodologically distinct educational assessments,

capturing both cognitive competencies and background variables from a wide range of sources (Fraillon et al., 2019; INVALSI, 2018).

From a methodological perspective, the study adopts a multilevel modeling framework to address the hierarchical nature of the data: students nested within schools (Goldstein et al., 2002; Enders & Tofighi, 2007). This structure violates the assumption of independence of observations in classical regression models, which justifies the use of multilevel techniques to partition variance across levels and obtain robust parameter estimates. The modeling proceeds sequentially: beginning with the empty model to estimate intra-class correlation coefficients (ICCs), and progressing through models that include student-level covariates (e.g., gender, socioeconomic status, digital self-efficacy) and school-level predictors (e.g., ICT infrastructure).

In line with best practices in multilevel analysis, continuous covariates, most notably the CIL scores and socioeconomic indices, are centered within clusters (CWC) to distinguish between within-school and between-school effects. This strategy ensures a more nuanced interpretation of fixed effects, isolating individual-level influences from contextual effects tied to school environments (Enders & Tofighi, 2007). Group-level means of these variables are then introduced at the second level, allowing us to separately model compositional and contextual relationships.

The digital competencies measured through ICILS are grounded in a robust framework that defines Computer and Information Literacy as the ability to use digital tools to gather, create, evaluate, and communicate information effectively (Fraillon et al., 2019; Vuorikari et al., 2022). The CIL scale is constructed using Item Response Theory (IRT) models, specifically the Rasch model (Rasch, 1960) for dichotomous items and the Partial Credit Model for polytomous responses (Masters & Wright, 1997), and scores are reported as plausible values. Each student is assigned five plausible values to represent their CIL proficiency, and all analyses involving these scores are repeated across these five draws, with final estimates pooled according to multiple imputation principles (Rubin, 2004; Adams et al., 1997).

The INVALSI assessments, on the other hand, are based on a national Rasch-calibrated scale with a mean of 200 and standard deviation of 40. For Grade 8, performance in Italian and Mathematics is reported across five proficiency levels, while English (Reading and Listening) results are categorized according to CEFR levels (INVALSI, 2018). By aligning both measurement systems on the basis of common psychometric foundations and shared population units, this study enables meaningful comparisons between digital and traditional literacies.

At the student level, variables related to ICT use inside and outside school and frequency of technology-based activities (Scherer et al., 2017; Van Deursen et al., 2016). At the school level, indicators of institutional digital capacity, such as the presence of ICT coordinators or digital teaching resources, are introduced as contextual predictors. These multilevel elements offer insights into whether the digital-educational link is merely correlational or embedded within broader educational structures (Tamborg et al., 2018; Livingstone et al., 2023).

Preliminary findings from the exploratory phase offer compelling insights. Specifically, correlation analyses conducted on the merged datasets reveal a strong and stable association between students' CIL scores and their performance across all four INVALSI subject areas for both years considered (2018 and 2023). In both years, Pearson correlation coefficients between the CIL scale and each of the INVALSI test scores ranged from approximately 0.56 in 2018 to over 0.60 in 2023, indicating a consistent and meaningful relationship between digital competencies and academic achievement.

This pattern of associations appears remarkably uniform across subjects, suggesting that digital literacy may not be uniquely tied to any single academic domain but instead supports general academic functioning, perhaps by enhancing students' capacity to access, process, and communicate information effectively across contexts. Notably, the slight increase in correlation strength in 2023, relative to 2018, may reflect the heightened integration of digital technologies in both instruction and assessment following the COVID-19 pandemic, which could have reinforced the role of digital competence as a cross-cutting academic enabler.

In addition to correlation analyses, early multilevel modeling results indicate that while school-level variance in INVALSI performance is generally modest, there are indications of stronger between-school effects in English subjects, particularly in 2018. This may reflect differences in instructional practices, resource allocation, or exposure to digital content in language learning contexts. These initial observations highlight the importance of accounting for institutional context when analyzing the digital–academic performance relationship.

Taken together, these preliminary findings support the broader hypothesis that digital competencies function as an academic amplifier, with implications that extend across subject boundaries and educational

settings. As digital environments become increasingly central to how students learn, these competencies may act not just as enablers of specific skills, but as foundational resources that facilitate deeper engagement with the curriculum as a whole.By combining rigorous psychometric methodology, sophisticated statistical modeling, and an integrated dataset spanning national and international sources, this study aims to make a meaningful contribution to educational research and policy. Our findings will have implications for the design of curricula, teacher training, and equity-focused interventions, particularly in contexts where digital access and skill levels remain unevenly distributed. The inclusion of socio-demographic controls also allows us to critically assess whether digital advantages are equitably distributed or reinforce existing educational disparities (Scheerder et al., 2017).
In conclusion, this research provides a novel, methodologically grounded exploration of how digital competencies, as measured by internationally benchmarked tools, relate to academic success in foundational school subjects. It emphasizes the value of cross-dataset linkage and multilevel modeling in capturing the complex interplay between individual skill, institutional context, and educational outcomes. As education systems continue to navigate digital transformation, understanding these dynamics will be critical to shaping inclusive, effective, and future-ready learning environments.

Keywords: DIGCOMP, Multilevel model, digital engagment, comparison 2018-2023 results

Automatic Item Generation through AI in Educational Testing: a Systematic Review

Francesca La Russa – Rita Marzoli – Antonella Mastrogiovanni – Alessia Mattei

Introduction
Artificial Intelligence (AI), a branch of computer science focused on creating machines capable of performing tasks that typically require human intelligence (Lu, 2019), has had a profound impact across various sectors, including educational assessment. Key applications of AI in this domain include automated scoring, computer-adaptive testing (CAT), and automatic item generation (AIG) (De Santis *et al*., 2025). AIG involves the use of algorithms to produce test items, based on predefined rules or models, that vary in content and difficulty while maintaining psychometric quality (Ahmed *et al*., 2024; Gierl *et al*., 2021). This approach allows for the efficient and scalable production of test items, significantly reducing the time and cost associated with manual item development.
As noted by Song & Du (2025), many studies have examined traditional AIG methods, such as rule or template based approaches, while fewer but more recent articles have focused on AIG through Generative AI (GAI) technologies, that create new content (text, images, audio, video, etc.) learning from patterns in existing data (Elmourabit *et al*. 2024), and Large Language Models (LLMs), a form of GAI generating human-like texts and performing complex language-related tasks with high precision.
Object and Aims
This study presents the results of a systematic review aimed at analyzing, synthesizing and assessing the quality of the existing literature on the use of AI for AIG. Specifically, it seeks to identify:

a) the types of items generated for student performance assessment;
b) the approaches and techniques used to generate these items;
c) whether, and in what ways, the quality of automatically generated items is evaluated.

Data and Methods
The systematic review follows the guidelines reported by Petticrew & Roberts (2006) in order to reduce bias and increase rigor and reproducibility. The systematic review follows a structured search strategy to locate and select relevant studies, employing predefined inclusion and exclusion criteria. The screening process is documented using the PRISMA Flow Diagram (Page *et a*l. 2020) A comprehensive literature search was conducted across four databases: APA PsycINFO®, Education Source Ultimate, ERIC, and Scopus. The Population, Concept, and Context (PCC) framework (Pollock *et al*. 2023) was adopted to develop the search string. The general search query included terms such as “automatic item generation”, “Artificial Intelligence”, “evaluation” or “assessment”, and “students”, which were refined for each database to optimize results.
After deduplication, the title and abstract of 291 records were independently screened by two reviewers in Rayyan, a web-based review management tool. The following inclusion criteria were applied:

1. Focus on automated item generation through AI

2. Application in testing contexts
3. Application in educational settings

No restrictions were imposed regarding study design, enabling the inclusion of theoretical, methodological, empirical, qualitative, and quantitative research.

Following the screening process and collaborative resolution of conflicts, 42 studies were selected for full-text analysis. During full-text screening, articles were excluded if any of the core concepts were addressed only superficially (e.g., mentioned in passing) or if the connections between AI, automatic item generation and educational assessment were insufficiently substantiated. After full-text screening, 9 articles were retained for coding and synthesis. For each article, citation information was exported into an Excel spreadsheet and key domains were abstracted based on the aims of the review (e.g. context of the study, type of automatically generated items, technical approaches to AIG; quality evaluation, outcomes and claims).

Finally, the GRADE-CERQual (Lewin *et al*. 2018), an approach for assessing how much confidence to place in the findings of a qualitative evidence synthesis, was employed to assess the methodological limitations, coherence, adequacy of data, and relevance of the included studies.

Results

The reviewed studies are all recent publications (from 2022 to 2025), underscoring the growing and timely interest in this field.

The findings of this systematic review reveal that most of the studies on AI-based AIG focus on the creation of multiple-choice questions (Elmourabit *et al*. 2024; Ahmed *et al*. 2024; Chan *et al*., 2025; Coskun *et al*., 2024; Kiyak *et al*., 2024; Marandi & Hosseini, 2024; Anggoro & Praitwi, 2023), with a smaller subset also including fill-in-the-blank items (Marandi & Hosseini, 2024) and constructed responses (Bulut & Yildirin Erbasli, 2022; Christ *et al*. 2024).

The application of AIG is primarily situated in higher education contexts, especially within STEM disciplines (Elmourabit *et al*. 2024; Ahmed *et al*. 2024; Chan *et al*., 2025; Coskun *et al*., 2024; Kiyak *et al*., 2024; Christ et al. 2024) and language learning (Marandi & Hosseini, 2024; Bulut & Yildirin Erbasli, 2022; Anggoro & Praitwi, 2023).

Most studies adopted approaches based LLMs, exploring a range of prompting strategies, from generic instructions to more detailed prompts that include source materials and examples (Chan *et al.*, 2024; Ahmed *et al*., 2024; Elmourabit *et al*., 2024). Coskun *et al*., 2024 combined LLMs with template-based methods, highlighting an emerging trend toward hybrid generation strategies.

In terms of quality assessment, analysis of students' performance (Elmourabit *et al*. 2024; Ahmed *et al*. 2024; Marandi & Hosseini, 2024) psychometric analyses (Coskun *et al*., 2024; Kiyak *et al*., 2024), expert reviews (Chan *et al*., 2025; Bulut & Yildirin Erbasli, 2022), and student feedback (Ahmed *et al*. 2024; Coskun *et al*., 2024) indicates that AI-generated items are generally valid, well-structured, and comparable in quality to human-authored items. Reported benefits include increased efficiency, reduced workload, and opportunities for personalization, while challenges relate to the continued need for manual review, concerns around transparency, and potential algorithmic bias.

The review identifies key emerging research trends on AI-based AIG, including the integration of adaptive learning technologies, advanced prompt engineering, and AI-human collaboration. Despite these promising directions, there remains a need for further research into the effectiveness of generating different item formats (e.g., open-ended question) across a variety of educational contexts and disciplines.

Finally, the application of the GRADE-CERQual framework to the included studies reveals varying levels of confidence in the reported findings. High confidence was attributed to findings from Chan *et al.* (2025); Elmourabit, *et al*. (2024), which were supported by rigorous methodologies, substantial data, and strong coherence with review objectives. Moderate confidence was assigned to findings from Ahmed *et al*. (2024); Coşkun *et al.* (2024); Marandi & Hosseini (2024); Kiyak e*t al.* (2024), which showed promising results but were constrained by reliance on expert review rather than formal testing and limited generalizability. In contrast, low confidence was assessed for findings from Bulut & Yildirin Erbasli (2022); Anggoro & Praitwi (2023); Christ *et al*. (2024)., primarily due to small-scale or informal evaluation designs.

Overall, the CERQual assessment highlights a need for more standardized empirical validation of AI-based AIG systems.

Keywords: Systematic review, Automated item generation, Artificial Intelligence, Educational Assessment

The M.I.E.L.E. Project: Digital Transition and Distributed Leadership in Educational Institutions

Monia Berghella – Paolo Davoli – Francesco Sacchetti

Introduction

School principals play a key role in supporting innovation in teaching methodologies (Paletta, 2020, 2023), by fostering professional learning communities that enhance student achievement, particularly through the integration of technology. Their leadership is also crucial in overseeing the digitalization of administrative and managerial processes essential to the strategic functioning of schools. In recent years, this role has been further supported by the resources introduced by Italy's National Recovery and Resilience Plan (PNRR), Mission 4, Component 1, aimed at driving digital transition in education.

To facilitate these processes, the Emilia-Romagna Regional School Office (USR-ER) has promoted the M.I.E.L.E. project—Management Innovation and Educational Leadership Evaluation—a three-year training and action-research initiative aimed at enhancing the professional development of school leaders and their communities. The project implemented peer-based and on-the-job training practices, engaging both principals and middle-management teams within their schools.

This paper focuses on one specific aspect of the project: data collected through the use of a Checklist (see Level-1 below), a self-assessment and guidance tool used to monitor digitalization processes within schools as a part of professional learning communities.

In 2025, the third year of the project emphasized digital transition processes promoted by the PNRR Education Program. It involved 151 school principals in the region, organized into 13 local groups facilitated by a peer principal, focusing on two thematic areas:

a) digital curriculum and methodological innovation;

b) digital management of strategic and administrative school projects.

The activities were structured on three levels:

- Level 0: Two peer focus groups among principals—at the beginning (to choose the thematic area) and at the end of the project.
- Level 1: At each participating school, principals conducted two focus groups with their local PNRR practice communities and selected teachers to explore the chosen theme using a digital checklist designed by the USR-ER.
- Level 2: Still to be implemented—local meetings of PNRR practice communities for further training and thematic deep-dives.

Objectives and Research Hypotheses

The ongoing study of the extensive documentation produced by participating principals and schools aims to examine in-service training practices and peer learning processes involving school leadership teams (both teaching and administrative). Findings will be published in a dedicated USR-ER volume (Desco et al., 2025).

The primary goal of the checklist (Level 1) was not merely data collection, but rather to stimulate reflection among school leaders and teaching communities on innovation processes fostered by the PNRR and to let each school to monitor current state of teacher collaboration and educational leadership.

The checklist was developed based on European frameworks—DigComp (2013–2022), DigCompEdu (2017), and DigCompOrg (2015)—and included 8 key areas:

1. Educational vision and digital transition
2. Curriculum change, digital curriculum, and teaching methods (aligned with Decrees 170/2022, 19/2024, and 65/2023)
3. Platforms and working practices:
   a. Administrative platforms and digital secretarial systems
   b. Digital gradebooks
   c. Learning Management Systems
4. Teacher collaboration and digital communication
5. School safety and data privacy
6. Professional development, strategic training, and educational research (aligned with Decree 66/2023)
7. Equipment, infrastructure, learning environments (aligned with School 4.0)

8. Areas for improvement

Data and Methodology

The digital checklist was completed by working groups in 115 of the 151 participating schools, following two focus groups held by each school's PNRR practice community (Level 1). The checklist contained:

- 224 Likert-scale items
- 33 binary (yes/no) questions
- 11 multiple-choice items
- 10 open-ended questions

Each participating school conducted two focus groups—before and after checklist completion—with an interval of 2–3 weeks. These allowed for a qualitative exploration of checklist themes and provided insight into processes of resistance or engagement with the digital transition.

The data analysis performed in this paper included descriptive statistics, correlation and multivariate analyses (also comparing regional and school-level differences), and the use of simple AI tools to analyze open responses.

Preliminary Results

Some initial findings:

1. **Digital Innovation and Strategic Orientation.** Over 90% of schools report that digitalization is at least partially integrated into their educational vision. There is general awareness of the importance of digital transformation, but the shift to systemic and structured practices is still limited. While many schools include digital elements in strategic documents (PTOF, RAV, PdM), only 32% have implemented a comprehensive monitoring system. The role of the Digital Animator is fairly common, but their connections to strategic governance (e.g., coordination with middle management or administrative managers) remain weak in 40% of schools.
2. **Infrastructure and Organizational Management.** This area showed the greatest variability. About 65% of schools report adequate technology provision, but ongoing maintenance remains a challenge. Only 25–30% of schools have stable technical staff for support; most rely on irregular or outsourced assistance. Digital operations are often reliant on "heroic individuals" or ad hoc solutions, indicating a lack of systemic oversight and dedicated professional resources.
3. **Distributed Leadership and Organizational Culture.** Leadership remains largely centralized in the role of the school principal, although some movement toward distributed leadership is emerging. Only 15–20% of schools have formal and active digital middle-management roles. Many refer instead to informal "working groups" or "thematic coordinators." Schools with established distributed governance tend to show better alignment between strategic vision and operational practices. Thus, while human capital is present, it needs to be invested in, recognized, and legitimized. Distributed leadership is fragile but promising.
4. **Methodological Innovation and Classroom Use of Digital Tools.** Around 70% of schools report adopting innovative digital teaching practices, but assessment of their impact is sporadic. Innovative learning environments have been primarily supported by PNRR funding (e.g., Next Generation Labs & Classrooms). Just over half of the schools have attempted to measure the impact of new environments on student learning. Digital use in classrooms is active, but often episodic. Stronger instructional leadership is needed to integrate innovation with curriculum design and student assessment.
5. **Overall Reflection.** Schools in Emilia-Romagna involved in the M.I.E.L.E.3 project show momentum, driven by strong visions and numerous micro-innovations. However, innovation is not yet systemic. Greater investments in human capital, distributed governance, sustainable infrastructure, and digital project management are needed.

Keywords: Digital competencies, distributed leadership, project and middle management, school principals

SESSION 3. INTERNATIONAL LARGE-SCALE ASSESSMENTS (ILSAS) METHODS AND RESULTS

ORGANIZER: UNIVERSITY OF BATH
COORDINATOR: ANDRES SANDOVAL HERNANDEZ
20TH NOVEMBER: 2.00 P.M. – 4.15 P.M. {AULA 5 – RESEARCH 9}

Italian students expected electoral participation: insights from IEA ICCS 2022

Michele Cardone – Ornella Papa – Rita Marzoli

Introduction

This study investigates the opinions, interests, and attitudes of Italian students in the eighth grade regarding relevant political and social issues, particularly focusing on expected electoral participation. The topic of abstentionism is currently of great interest given the low electoral participation observed especially among younger people, in our country as well as in others (Smets 2010). Previous studies have verified that the prediction of political participation is positively related to actual participation in adulthood (Keating, Green & Janmaat, 2015). (Keating and Janmaat, 2016). Moreover, although the political socialization of adolescents is considered crucial for their future political participation, there are not many in-depth studies conducted on this topic at the national and international levels (Myoung & Liou, 2022). The data analyzed relate to the IEA survey (International Association for the Evaluation of Educational Achievement) ICCS (International Civic and Citizenship Education Study) of 2022. The Survey collects data on citizenship education and the civic knowledge of students around 13 and a half years old, as well as their opinions and attitudes towards relevant political and social issues (Schulz et al., 2023). The administration, in Italy carried out by INVALSI, includes a Cognitive Test on civic and citizenship knowledge, as well as a Student Questionnaire that collects their characteristics, opinions, interests, and attitudes in political and social domains, for example the intention to participate in voting. ICCS 2022 is the third Cycle of the Survey (following ICCS 2009 and ICCS 2016) involving 22 Educational Systems worldwide, with an increase in participation from European countries compared to previous editions. The European Union, in fact, recognizes active citizenship as a key competence and recommends educational paths for its promotion and the monitoring of the levels achieved. For the countries that have participated in all editions, such as Italy, the data are comparable to those of previous editions, allowing for the tracking of the evolution over time of the results achieved.

Subject, objectives and research hypotheses

This study focuses on Italian students' predictions of participating in voting, analyzed in relation to the characteristics and opinions on political and social issues of the students. The data used are mainly those of the Student Questionnaire, but the results of the Cognitive Test relating to civic and citizenship knowledge, are also considered. The analyses include comparisons with international data and with those of previous Cycles. The sample consists of 2,945 students, attending eighth grade in 154 schools distributed across the national territory (Invalsi, 2023), who completed both the Student Questionnaire and the Cognitive Test. The objective is to contribute to the identification of the motivations behind the growing youth abstentionism. It seems important to describe the characteristics of the students who are less likely to vote: a first step to understanding a complex and widespread phenomenon.

Method

The data are analyzed in a comparative perspective both with previous Cycles, particularly with the most recent one of 2016, and with international data. Through correlations, the association of students expected electoral participation is assessed with declared political and social opinions and interests; also, the association with civic and citizenship knowledge, and with some variables related to socioeconomic and cultural background is verified. The data analysis is carried out using IDB Analyzer, specific software created by the IEA.

Results

The results obtained show that the factors most associated with the intention to vote are: Students' beliefs about the importance of conventional citizenship activities (0.39), Students' expected participation in environmental protection activities (0.39) Students' expected participation in legal activities (0.37), and

scores on the Cognitive Test (0.37), higher on average for females; a strong association is also found with the variable relating to self-efficacy as "citizens" (0.35). A moderate positive association is found with the intention to continue studying - Expected ISCED (0.29), with Highest parent's interest in politics (0.28), and with home literacy resources (0.27). The students' expected electoral participation is negatively associated with dissatisfaction and distrust towards the political system (-0.22); Italian students are more dissatisfied and distrustful towards the political system and institutions, particularly the Parliament, compared to students in other countries. In fact, 56% of Italian students believe that their country's political system does not work well (compared to 45% of the international average); 64% of Italian students believe that parliamentarians do not adequately represent the interests of young people (compared to 56% of the international average) and 70% of our students believe that political representatives do not care enough about people's wishes (compared to 62% of the international average). On the contrary, interest in political and social issues shows a higher percentage among Italian students (39%) than among students at an international level (30%). Compared to the 2016 Cycle, trust in Parliament in Italy has decreased by 10 percentage points.

Conclusions

In accordance with the results of previous studies on the importance of citizenship education in the (Keating, Green & Janmaat, 2015; Keating & Janmaat, 2016), this study confirms its great importance also for the purposes of electoral participation: students with greater Test scores are more likely to vote once they reach adulthood. However, more and more students plan to abstain from voting due to dissatisfaction and distrust towards governments institutions rather than disinterest and disengagement. Believing that politics does not sufficiently address the needs of citizens and issues particularly dear to young people, such as the environment and civil rights, they turn away from it, opting for new forms of activism. The attitude towards politics is therefore critical rather than apathetic or disinterested, as one might imagine (Keating, Green & Janmaat, 2015). On the contrary, Italian students are more interested in political and social issues compared to the ICCS average and discuss them more with their parents. Already in the ICCS 2016 Cycle, although their trust in institutions had decreased, Italian students reported talking with their parents about socio-political issues more frequently than in the ICCS 2009 Cycle (Malak-Minkiewicz & Torney-Purta, 2021). Understanding what students think about political issues is important to grasp the reasons behind the growing youth abstentionism, which has long reached 50% among non-graduates (Tuorto, 2018). Abstentionism is a phenomenon that, if not understood and contained, can interfere with an inclusive and representative democracy. Already during the political elections of 2022, the overall abstention rate in Italy reached 36%, placing us at the bottom among Western European countries in terms of voter turnout (Emanuele, V. & Marino, B., 2022).

Keywords: Electoral participation, students' opinion, political and social issues, Civic and Citizenship knowledge, IEA ICCS 2022.

Beyond Achievement Gaps: Gender Differences in Educational Quality Configurations Across OECD Countries through Latent Profile Analysis

Manuel Cheyre

Educational quality has emerged as a central organizing principle in global education policy, featured prominently in international frameworks such as the UN Sustainable Development Goals and driving reforms across educational systems worldwide. Yet beneath this apparent consensus lies a fundamental paradox: while there is widespread agreement on the importance of educational quality, the concept itself remains remarkably ambiguous, operating as what Clarke (2014) describes as an "empty signifier" onto which diverse and often contradictory meanings are projected. This conceptual ambiguity, rather than fostering diversity in educational evaluation, has facilitated the consolidation of a dominant vision that reduces educational quality almost exclusively to academic achievement measured through standardized assessments.

This study challenges this dominant tradition by reconceptualizing educational quality not as an intrinsic property of educational systems, but as an evaluative judgment that can be constructed from multiple legitimate axiological frameworks. Drawing on evaluation theory (Popham, 1993; Scriven, 1995; Tyler, 1942, 1991), I argue that quality emerges from systematic judgments about educational phenomena based

on explicit components and standards that reflect specific purposes and values about what education should accomplish. This theoretical shift opens space for recognizing that different axiological frameworks—different conceptions about the fundamental purposes of education—may generate different but equally valid conceptualizations of what constitutes "good education."

Using latent profile analysis (LPA) on 2022 PISA data from OECD countries, this study examines whether educational quality manifests differently across gender and countries by analyzing multiple components of students' educational experience beyond academic achievement. The analysis incorporates five key dimensions: academic performance (mathematics, reading, and science), student-teacher relationships, socialization experiences, academic self-efficacy, and student well-being. These components reflect a more comprehensive understanding of educational experience that aligns with broader theoretical frameworks about educational purposes, including Biesta's (2009) tripartite model of qualification, socialization, and subjectification, and Brighouse et al.'s (2018) educational goods framework.

The focus on gender differences is theoretically motivated by the recognition that different groups may experience education differently and that these differences may reflect distinct pathways to educational value rather than deficits to be corrected. If educational quality configurations vary systematically between male and female students, and if these patterns differ across countries, this would provide empirical evidence for the theoretical proposition that quality is not a universal, monolithic concept but rather can take multiple legitimate forms depending on how different dimensions of educational experience are experienced, valued, and combined.

Preliminary analyses reveal distinct educational quality profiles that differ between male and female students between OECD countries. These findings go beyond simply documenting achievement gaps to reveal that male and female students are not only receiving different educational experiences but that these experiences can be interpreted and valued differently depending on one's axiological framework about educational purposes (Biesta, 2015a). For instance, profiles characterized by high academic achievement but lower socialization scores might be valued differently by frameworks that prioritize economic productivity versus those that emphasize democratic participation or social cohesion. Results also suggest that male-female differences exhibit diverse patterns across different cultures and educational systems, indicating that different contexts—potentially guided by distinct axiological frameworks—may interpret different configurations as representing “quality education” all the same.

Importantly, these findings should not be interpreted as suggesting that males and females require different types of education or that observed differences are natural or desirable. Rather, they demonstrate that the current "one-size-fits-all" approach to defining and measuring educational quality may be insufficient for capturing the complexity of educational experience and the diversity of ways that education can be valuable. The existence of different quality profiles suggests that educational systems should develop more nuanced approaches to quality evaluation—approaches that recognize diverse pathways to what might be considered "good education" rather than relying exclusively on singular achievement metrics.

These results have significant implications for both educational research and policy. Methodologically, they demonstrate the feasibility of developing more comprehensive approaches to educational quality that can accommodate multiple dimensions without privileging any single component a priori. The latent profile approach reveals naturally occurring configurations of educational components that emerge from the data rather than being imposed by researchers, providing empirical evidence for theoretical arguments about the multidimensional nature of educational quality.

From a policy perspective, these findings challenge the hegemony of test-based accountability systems (Verger et al., 2019) and suggest the need for evaluation frameworks that can recognize and value different forms of educational excellence (Appels et al., 2024). Rather than asking whether one profile is "better" than another, the question becomes which profiles are most aligned with explicit educational purposes and values in specific contexts (Biesta, 2015b; Brighouse et al., 2018). This shift from universal standards to purpose-driven evaluation could enable more democratic and inclusive discussions about educational priorities.

The study also highlights the importance of making axiological frameworks explicit when using large-scale assessment data. PISA and similar assessments are often treated as objective measures of educational quality, but this research demonstrates that the same data can be interpreted differently depending on what one values in education. When policymakers use PISA scores to make judgments about educational quality, they are implicitly privileging certain educational purposes (typically economic competitiveness) over others (such as social cohesion, democratic participation, or individual flourishing).

Furthermore, this research contributes to ongoing debates about gender and education by shifting the focus from achievement gaps to quality configurations. While gender achievement gaps remain important, this study suggests that focusing exclusively on academic performance may miss other important dimensions of how males and females experience education. Some of the identified profiles might represent different but equally valuable forms of educational engagement that current evaluation systems fail to recognize or appreciate.

The theoretical framework developed in this study also has broader applications beyond gender differences. The same analytical approach could be used to examine how educational quality manifests across other demographic groups, cultural contexts, or educational systems. This could provide empirical evidence for the theoretical argument that educational quality is inherently contextual and that effective evaluation systems must be capable of recognizing diverse forms of educational value.

Looking forward, this research opens several important avenues for future investigation. First, longitudinal studies could examine how different quality profiles relate to various life outcomes, potentially revealing that different configurations lead to different types of success and fulfillment. Second, qualitative research could deepen understanding of how students, teachers, and families experience and interpret different quality configurations.

In sum, this study provides empirical evidence for the theoretical proposition that educational quality is not a monolithic concept but can take multiple legitimate forms. By demonstrating that educational quality profiles vary systematically across gender, it challenges the universality of current quality frameworks and calls for more nuanced, purpose-driven approaches to educational evaluation. Most importantly, it suggests that the question should not be "what is educational quality?" but rather "what forms can educational quality take when serving different purposes and values?" This shift in perspective could enable more democratic, inclusive, and ultimately more just approaches to improving educational systems worldwide. By making visible the axiological choices embedded in quality evaluations, this research contributes to more informed and transparent discussions about what we collectively value in education and how our evaluation systems can better serve those values.

Keywords: Quality education, comparative education, axiological frameworks, multidimensional assessment

Analyzing the Use of Accommodations and Exclusion Rates for ICILS 2023: The French Example

Elodie Vezon Persem – Adrien Fernandez

This study explores the effect of accommodations on the exclusion rates for the International Computer and Information Literacy Study (ICILS) 2023, focusing on France as a case study. In the context of the increasing inclusion of students with special educational needs, the study investigates whether the new accommodations implemented for ICILS 2023 have contributed to reducing exclusion rates in France. By examining the use of these accommodations and their impact on student participation, this research connects to IEA studies, specifically focusing on the methodological challenges of standardized assessments for students with special educational needs. The findings indicate a limited use of accommodations and suggest that, while some students benefited from adjustments, these accommodations did not significantly reduce exclusion rates. The key conclusion calls for a methodological reevaluation to better align international assessment practices with inclusive educational paradigms.

Introduction

The international ICILS study targets a representative sample of 8th-grade students from participating countries (Fraillon, J. Ed., 2024). In France, the student population is diverse, including those with disabilities, linguistic differences, and other educational needs (RERS 2023). French law, particularly the 2005 Disability Act and subsequent legislation, mandates the inclusion of students with special educational needs in mainstream schools (LOI n° 2005-102). The question remains whether the accommodations provided in large-scale assessments, such as ICILS 2023, are sufficient to reduce exclusion rates and allow these students to participate.

This research is significant because it addresses the methodological challenges of including diverse students in standardized assessments. While inclusion is a well-established educational paradigm in France and internationally (UNESCO, 1960, 1994), its application in large-scale assessments such as ICILS remains a

point of tension especially compared to national large-scale assessments (Hick, M., et al.,2024). This study seeks to explore whether the new accommodations introduced for ICILS 2023 can help bridge the gap between educational inclusivity and standardized testing.

Hypothesis: The introduction of new accommodations has allowed more students to participate and reduced the exclusion rate within schools.

Data Sources, Methods, and Statistical Techniques

This study uses secondary data from ICILS 2023, focusing specifically on France. The data includes information on student participation, the accommodations provided, and the rates of exclusion due to educational needs. The primary method involves descriptive analysis of the use of accommodations, such as extended time, separate rooms or assistance in reading instructions. Statistical techniques, including frequency analysis and comparison of exclusion rates before and after the introduction of accommodations, are used to assess the impact of these accommodations on participation.

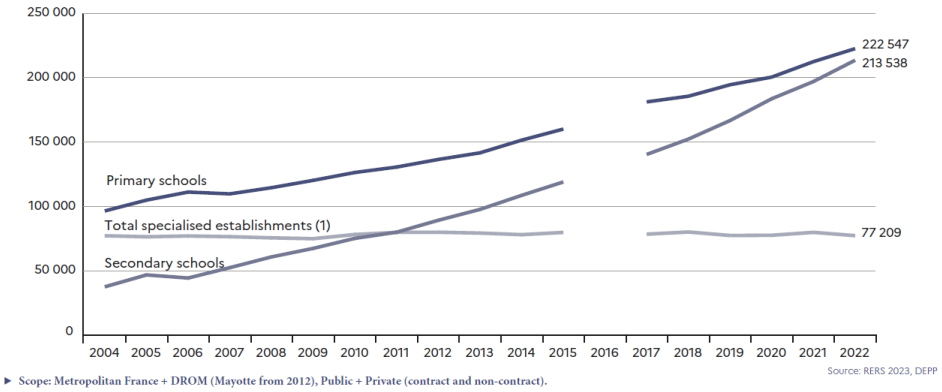
Results and Discussion

The introduction of accommodations for students with special educational needs (SEN) in the ICILS 2023 assessment represents a significant step towards improving inclusivity in international large-scale assessments. However, despite the availability of these accommodations, their uptake in France has been limited, raising important questions about their effectiveness in reducing exclusion rates. This section presents an analysis of the usage of accommodations in France, their impact on exclusion rates, and the implications for future assessments.

Understanding the Population in the Inclusive French Educational System

Inclusion has become a pivotal theme in education worldwide, particularly since the 1989 UN Convention on the Rights of the Child and the Salamanca Declaration of 1994, which emphasized the need for inclusive education for all children, regardless of their abilities or backgrounds. This was a turning point in educational history, as systems worldwide began to shift from segregated educational models to ones that prioritize inclusivity. In this context, the 2030 Agenda for Sustainable Development, specifically its fourth goal, further reinforces the drive for inclusive education, advocating for equal access to quality education for all. France has strongly embraced this inclusive approach, with legislation such as the law of February 11, 2005, which established a framework for the inclusion of students with disabilities in general education settings. This has led to a significant increase in the number of students with disabilities included into mainstream classrooms. The number of students with disabilities in these schools has tripled, rising from 155,361 at the start of the 2006 school year to 436,085 in 2022. By September 2023, they represented 3.6% of all students in French schools (Figure 1) and by September 2024, 4.36 %. Similarly, at the start of the 2021 school year, students with a non-native language background numbered 77,400 (across primary and secondary levels), reflecting a 20% increase from the previous year. Additionally, we must take into account and add students with less severe disorders, students who are ill or have suffered an accident, those with language and learning difficulties and undiagnosed students. While there are no precise studies on the number of students with these conditions, the prevalence of such disorders may provide some insight. For example, in France, approximately 8% of pupils suffer from specific language and learning disorders. As a result, the population of students in France is highly heterogeneous, with a wide range of educational needs. These students are an integral part of the reference population for the sample of ICILS 2023.

1. Figure 1: Schooling Trends for Children and Adolescents with Disabilities



▶ Scope: Metropolitan France + DROM (Mayotte from 2012), Public + Private (contract and non-contract).

1. Excluding children attending school for short periods.
2. Data missing due to an administrative strike by referral teachers.

Note: numbers include pupils in shared schooling, who account for 3% of total enrolment.

This has consequences for the methodological aspects of deploying large-scale standardized tests. Indeed, some of these students have specific needs and benefit from adaptations and/or accommodations during their schooling. However, by definition, a standardized assessment provides identical content and testing conditions for the entire sample population.

If adaptations and/or accommodations are not available or integrated into the test during the administration of a large-scale standardized assessment, the student may be hindered in completing their work and therefore may not be assessed to their full potential. The first risk, therefore, concerns the validity of the psychometric results obtained for these students. Students with special educational needs may also be unable to complete the test at all. The second risk is the potential increase in the number of students excluded from the test. This then presents a significant issue for France: failing to meet the legal obligation of inclusion. Moreover, an increase in these exclusions could heighten the risk of inadequate coverage of our reference population. Thus, in France, there is a tension between the design of a standardized test and the target population of the test, which is, by definition, highly heterogeneous.

ICILS 2023: Use of Accommodations in France and Exclusion Rates

Until 2018, the IEA did not offer adaptations or accommodations for students with special educational needs. Three exclusion criteria were proposed to allow certain students with special educational needs not to participate in the ICILS test:

1 = Students with functional disabilities;
2 = Students with intellectual disabilities;
3 = Non-native language speakers.

ICILS 2023 introduced a set of accommodations to improve accessibility for students with educational needs. For the first time, the IEA (International Association for the Evaluation of Educational Achievement) allowed for adjustments such as the use of extra time, the provision of separate rooms, assistance from a reader or scribe, the option to disable the timer, additional break time between modules, and the possibility of providing an additional session on a second day. In France, these accommodations have generally been little used despite the characteristics of the population (Table 1). The most commonly used accommodation was the deactivation of the timer, which aligned with a typical accommodation used in France's national assessments. Only a small proportion of students, about 1.1%, benefited from this adjustment. Other accommodations, such as providing a reader or a scribe, were used even less frequently while this is very common in class.

Table 1: Percentage of Students who Received Accommodations for the ICILS 2023 Assessment

Accommodations	France (%)	International (%)
Additional break time between modules	0.1	0.2
Disable timer	1.1	2.3
3rd party screen magnifiers	0.0	0.0
Person to read materials	0.4	1.2
Additional session on second day	0.0	0.1
Separate room	0.1	0.5
Person to write responses	0.1	0.2

In France, 1.4% of students benefited from at least one type of accommodation, compared to an international average of 3.0%. 81.2% of students who received at least one type of accommodation in France received only one, compared to 68.9% on average internationally (Table 2). The low usage of accommodations, particularly in light of the growing number of students with SEN, may suggest a potential gap between the accommodations provided and the actual needs of the student population.

Table 2: Percentage of Students who Received at least one Type of Accommodation

	% of students
France	1,4
International	3,0

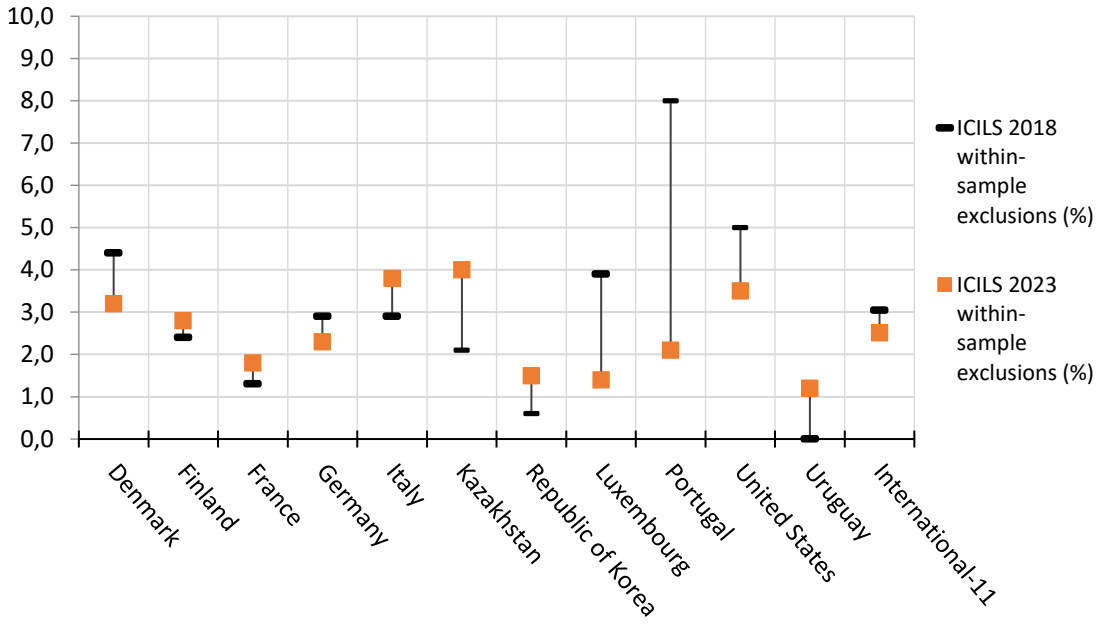
Table 3: Distribution of Students who Received at least one Type of Accommodation

	1 accommodation (%)	2 accommodations (%)	3 accommodations (%)	4 accommodations (%)	5 accommodations (%)	6 accommodations (%)
France	81,2	17,1	1,7	0,0	0,0	0,0
International	68,9	20,0	8,6	2,4	0,1	0,0

Did these Accommodations have a Positive Effect on the Exclusion Rate in France?

For ICILS 2023, 1.8% of the sampled students in France, enrolled in participating schools, were excluded from the survey, which is below the international average. This rate was 1.3% for ICILS 2018 (Figure 2). The same can be observed if we consider only the first two exclusion rates, leaving aside the exclusion of non-native language speakers because the accommodations were not designed to meet their needs. In 2023, the exclusion rate in France was 1.1%, compared to 1.6% internationally. Although these accommodations aimed to make the assessment more accessible, the exclusion rates did not drop but continued to rise. It is therefore assumed that the accommodations proposed by ICILS 2023, which, by definition, only address the administration modalities, only partially meet the needs.

Figure 2: Percentages of Students Excluded from the ICILS Target Population



In France, large-scale standardized assessments are conducted for all students every September from grade 1 to grade 10. The participation rate of students within participating schools is on average 93% in 2023 for grade 8. There are no exclusion codes that allow us to more specifically identify the reasons for non-participation. The accommodations offered are the same as those typically available to students in the classroom. These are maintained during the assessments. The accommodations proposed by ICILS 2023 are part of the accommodations provided at the national level, although other accommodations supplement the

list (such as workstation adjustments, the use of regular digital tools like text-to-speech software, etc.). Furthermore, the contents for these large-scale standardized assessments are designed to be accessible considering the population and our regulatory obligations (Universal Design for Learning approach, adaptations in braille and French sign language for sensory disabilities). The adaptation and accessibility of the test's evaluation tools are integral to the methodology. However, work is still ongoing with the provider of the digital testing platform to ensure that it is fully accessible and meets the WCAG criteria.
Two aspects seem essential to consider when addressing special educational needs: test accessibility (design of the evaluation tools, test design, accessibility of materials in paper or digital formats) and accommodations (administration modalities).
This reflection leads us to ask how to reduce exclusion rates in large-scale standardized assessments, whether national or international. How can exclusion rates in standardized assessments be reduced to ensure equitable participation for all students? How can equity be restored among students to produce reliable and valid statistical measures? How can we best represent the population within each country's education system? How can we integrate an inclusive approach into the methodology used in large-scale standardized assessments?
We suggest that methodological reflection be carried out along two lines:

- Defining the actual target population from which the samples are drawn: excluding a portion of students from the education system in the assessments may undermine the validity of the psychometric results.
- Designing evaluation tools within a universal and accessible approach.

By integrating these principles into the methodology of national or international assessments, it would be possible to reduce exclusion rates, enhance the equity of participation, and ensure more reliable and valid statistical measures, ultimately providing a fairer comparison between students from diverse backgrounds worldwide.

Conclusions

The introduction of accommodations in ICILS 2023 represents a positive step towards inclusivity, but the impact on exclusion rates has been minimal. This study suggests that the accommodations provided in ICILS 2023 are not sufficient to ensure equitable participation for all students, particularly in countries like France with a diverse student population. A methodological shift is required to align large-scale assessments with inclusive education policies. This could involve redefining the target population to more accurately reflect the diversity of students in participating countries and incorporating a more comprehensive set of accommodations into the design of the assessment tools.
Future research should explore how to better integrate inclusive design principles into the methodology of large-scale assessments, ensuring that all students, regardless of their needs, can participate meaningfully and have their educational progress accurately measured.

Keywords: Accommodations, exclusion rates, inclusive education, standardized assessment

Indicators for sustainability and climate change education in international large-scale assessments: early insights from the IMP>ACT Project

Riccardo Pietracci – Adam Coates – Maria Magdalena Isac – Laura Palmerio

Background and rationale

In recent years, international large-scale assessments (ILSAs), such as PISA, TIMSS, and ICCS, have increasingly incorporated elements related to sustainability and climate change education (SCCE) into their conceptual frameworks. This growing attention reflects the global recognition of the relevance of Education for Sustainable Development (ESD) within current and future educational agendas (Borgonovi et al., 2022; White et al., 2023; Reynolds and Komakhidze, 2022; Schulz et al., 2023). This data offers a valuable cross-national perspective not only on students' cognitive outcomes but also on how different educational systems approach ESD in terms of educational goals, cognitive and behavioral outcomes, teaching methods, and the challenges teachers face in addressing sustainability topics and issues.

The IMP>ACT for Action project

Launched in early 2024 and set to continue until the end of 2027, the IMP>ACT project aims to develop a comprehensive framework for assessing the impact of sustainability and climate change education. Funded

under the Horizon Europe programme, the initiative responds to a growing demand among policymakers and practitioners across Europe for robust, context-sensitive tools and approaches that support the evaluation of educational impact in this field.

The project is carried out by a consortium of five universities and seven societal partners, who have jointly contributed to discussions on the development of the assessment framework, ensuring both methodological soundness and relevance to diverse educational contexts and potential user profiles. Among its key objectives is the mapping of existing evaluation methods, monitoring strategies, and indicators related to SCCE at national and international levels. Within this framework, ILSAs are explored as a significant source of comparative and scalable data that can inform our understanding of how SCCE is addressed and measured globally.

The IMP>ACT for Action framework

At the core of the IMP>ACT project is the development of the IMP>ACT for Action framework, which places particular emphasis on action-oriented pedagogy and the assessment of competencies related to student agency and engagement. This approach values not only final outcomes, but also the processes involved in teaching and learning for sustainability. These include the quality of interactions between students and teachers, the integration of sustainability themes into project-based or inquiry-based learning, and the degree to which students are involved in decision-making processes or engaged in community-relevant actions.

A central tenet of the framework is its commitment to a pluralistic view of ESD. Rather than transmitting pre-defined knowledge, this perspective encourages students to engage with sustainability issues based on their own motivations and interests. Such an approach promotes active participation and strengthens students' sense of ownership and responsibility toward sustainability challenges. In addition, the framework is designed to capture the holistic nature of sustainability, combining scientific knowledge with economic, political, and social dimensions. This integrated perspective supports a deeper understanding of the interdependencies that characterize sustainable development and equips learners to navigate complex global issues.

Mapping SCCE indicators in ILSAs

Mapping study frameworks and questions used in ILSAs can enhance understanding of ESD implementation and its impact on student learning, allowing for a clearer picture of the educational processes and environments contributing to sustainability education (UNESCO, 2020). In particular, mapping ILSA scales and questions onto the IMP>ACT action framework can highlight the importance of action competencies as an outcome of ESD; support future research which wishes to use this framework; and identify concepts which are not currently covered by ILSA questionnaires.

Findings and implications

This work aims to provide a comprehensive overview of the current state of information on SCCE as captured by ILSAs such as ICCS, TIMSS, and PISA. Using the conceptual framework developed within the IMP>ACT project, the mapping classifies items across the dimensions of context, input, processes, and outputs (Sinakou et al., 2019).

From this perspective, relevant data from these ILSAs have been collected and analyzed, with existing items and instruments classified according to the identified dimensions. The analysis highlights both overlaps and gaps among frameworks and evaluates the comparability and consistency of the data. Particular attention has been paid to underrepresented aspects, such as process- and context-related indicators.

In particular, the findings show that most items used in international assessments are primarily input- and output-oriented, and are scattered across different frameworks. These items focus on quantitative aspects from a comparative perspective but may only partially capture the full range of factors influencing ESD effectiveness. Moreover, the contextual factors shaping how ESD is taught and learned, such as cultural values or local environmental challenges, are rarely captured, as international surveys must preserve comparability while balancing national specificities with relevance across all educational contexts.

Based on this mapping, preliminary recommendations have been formulated to enhance the scope and relevance of the contents of existing international tools. These recommendations aim to better capture the impact of educational practices in SCCE, paving the way for the development of international reference indicators that address not only inputs and outputs, but also the pedagogical processes and contextual factors that shape education for sustainability and climate change.

Keywords: Sustainability and Climate Change Education, International Large-Scale Assessments, Education for Sustainable Development, Assessment Frameworks

Trends in Socioeconomic Inequalities in Reading Literacy Across Grades in European Countries Over Time

Andrea Pietrolucci – Nathalie Vigna – Giuliana Parente

Introduction

Since Coleman's report (1966), sociologists of education have extensively studied the effect of socioeconomic status (SES) on students' achievement. Although SES can be defined in various ways, it is generally understood as a composite measure of an individual's economic and social position, commonly based on three key indicators: education, income, and occupation (Baker, 2014). Accordingly, parental resources in terms of educational attainment, occupational status, and income are widely recognized as central components of a student's SES. Achievement—typically measured through standardized test scores in subjects such as reading, math, and science—and educational attainment—reflected in the acquisition of degrees and formal credentials—are both considered key pathways to social status and mobility (Blau & Duncan, 1967; Sewell & Hauser, 1972). Children from low-SES families consistently underperform compared to their high-SES peers (Graetz, 1995), which can negatively affect their future academic and occupational trajectories.

Previous research has documented that socioeconomic disparities in academic achievement emerge early and tend to evolve over the educational career. Several studies have shown that gaps in reading and math performance are already visible in primary school and may either widen or stabilize in later stages, depending on institutional and policy contexts (Sirin, 2005; Reardon, 2011; van de Werfhorst, 2017). Some evidence suggests the relationship between SES and academic performance may vary across grades: while early stages of education often display wider SES-related gaps, the effect of family background may attenuate in later grades due to compensatory school experiences or student selection processes (Marks & Pokropek, 2019; Peng et al., 2019; Cilasun, 2013). These patterns are not uniform across countries, however, and may reflect structural differences in how education systems are organized Kim (2019). Empirical studies on this topic have also shown that there has been substantial stability (Hanushek et al., 2019) or even an increase (Chmielewski, 2019) in socioeconomic inequalities in academic performance over the last half century.

Although students' achievement can be investigated across various subjects, we consider reading literacy a foundational skill that plays a critical role in students' overall educational trajectories and life outcomes. In fact, proficiency in reading is not only essential for academic success across disciplines, but it also influences later outcomes such as access to higher education, employability, income levels, and civic engagement (Rodríguez et al., 2020). Inadequate literacy skills are strongly associated with early school leaving, social exclusion, and limited opportunities for lifelong learning. As such, fostering reading literacy from an early age is considered a central objective of education systems and a key lever for reducing social inequalities.

Building on these considerations and acknowledging that both grade level and contextual factors can moderate SES effects, this study provides a comprehensive examination of how socioeconomic inequalities in reading literacy vary across school grades, countries, and time. The main objective of this study is to assess the extent and evolution of socioeconomic inequalities in reading literacy, focusing on the role of parental education as a key indicator of SES. Adopting a comparative and longitudinal approach, the study investigates how these inequalities unfold across different educational stages and European contexts over time, using harmonized data from international large-scale assessments. While prior research has often focused on a specific stages of schooling or single countries, this paper adopts a comparative perspective based on harmonized data from two major international large-scale assessments: PIRLS, which focuses on the fourth grade, and PISA, which covers 15-year-old students. By linking multiple waves of both datasets from 2006 to 2022, the study traces the dynamics of SES-based inequalities over time and across educational transitions. The focus on European countries is particularly relevant, as they share common objectives in terms of equity and quality under the European Education Area, yet differ in their welfare regimes, levels of socioeconomic inequality, and institutional approaches to educational equity.

Data and methods

We integrate data from two international student achievement surveys: the Progress in International Reading Literacy Study (PIRLS) and the Programme for International Student Assessment (PISA). PIRLS evaluates reading comprehension among fourth-grade students in nearly 60 countries and is administered every five years, beginning in 2001. PISA evaluates students' achievement in various subjects, including literacy, mathematics, and science, among tenth-graders in 80 countries every three years since 2000. To

improve the comparability of the surveys, we include in our analytical sample all European countries observed in both PISA and PIRLS that have participated in at least three waves of the surveys starting in 2006. We do not use waves prior to 2006 because the number of participating countries in those years was much lower. The resulting analytical sample comprises nearly 1.2 million observations from 19 European countries. It is composed of 352,693 observations in four waves (2006, 2011, 2016, and 2021) from PIRLS and 841,653 observations in six waves (2006, 2009, 2012, 2015, 2018, and 2022) from PISA.

To investigate trends in socioeconomic inequalities in reading literacy, we use the standardized reading competence test scores administered by PIRLS and PISA as the dependent variable. In the current version of our work, we study the correlation between SES origins and the average test score (standardized by year and country to allow for cross-year and cross-survey comparisons) among three groups:

a. The average test score.

b. The low performers: a dummy variable specifying whether a student scores equal to or lower than the bottom 25% of the country-year distribution.

c. The high-performers: a dummy variable specifying whether a student scores equal to or higher than the top 10'% of the country-year distribution.

Future extensions of this work will also include measures of score dispersion, such as standard deviation or interquartile range.

We use parental education as a measure of socioeconomic background, operationalized as 1 if one parent has a tertiary degree and 0 if neither parent has one. In future versions of the work, we will extend this definition by adopting a more refined definition of parental education and including other dimensions of socioeconomic status, such as parental occupation, and material and cultural resources.

To describe how the association between parental education and reading literacy performance varies over time, we use a two-step analytical strategy. In the first step, we regress test score measures on parental education, controlling for gender and language spoken at home, for each survey wave and country separately. In the second step, we apply a meta-regression framework by using the resulting coefficients as a dependent variable (Stanley & Doucouliagos, 2012). More specifically, to describe the levels of socioeconomic inequalities in different years, we report the average values of the beta coefficients from Step 1. These values are weighted by the standard errors to account for differences in the accuracy of the estimates. The results are presented separately by wave and survey. Next, to examine variation in socioeconomic inequalities over time, we apply multilevel meta-regression with country fixed effects, specifying the survey waves as the independent variable.

Preliminary results

Preliminary results show large inequalities in literacy skills based on parental educational background (Figure 1). Indeed, over the waves considered, students from families with highly educated parents have an advantage of around 0.6 standard deviation (SD) in PIRLS and about 0.4 SD in PISA. Interestingly, these preliminary results show that this (relative) gap is larger in the early stages of education (fourth grade) and decreases slightly in later stages (tenth grade). However, we acknowledge that the smaller relative gap might be due to the fact that students from a more disadvantaged background are selected in higher grades, as lower-performing students from low-SES backgrounds are at a higher risk of dropping out. The same pattern applies to the likelihood of having a score in the bottom 25% or top 90% of the test score distribution.

When modeling the variation over time of inequalities in reading literacy attainment, different patterns emerge across the PIRLS and PISA surveys. Regarding PIRLS, inequalities in average achievement remain fairly stable over time. Conversely, inequalities in the likelihood of being a low performer increase by 1 percentage point over years, while inequalities at the top of the score distribution decrease by 2-3 percentage points. In the case of PISA, however, the pattern is more consistent, though it shows non-linear variation over time (Figure 2). Up until 2015, there we notice an increase in inequalities, which is particularly marked for average reading literacy achievement. Conversely, in 2018, inequalities in all three outcomes are likely to reduce. Finally, in 2022, the post-Covid wave shows no significant variation compared to 2006.

Figure 1. Average levels of inequalities in reading literacy achievement by year and survey.

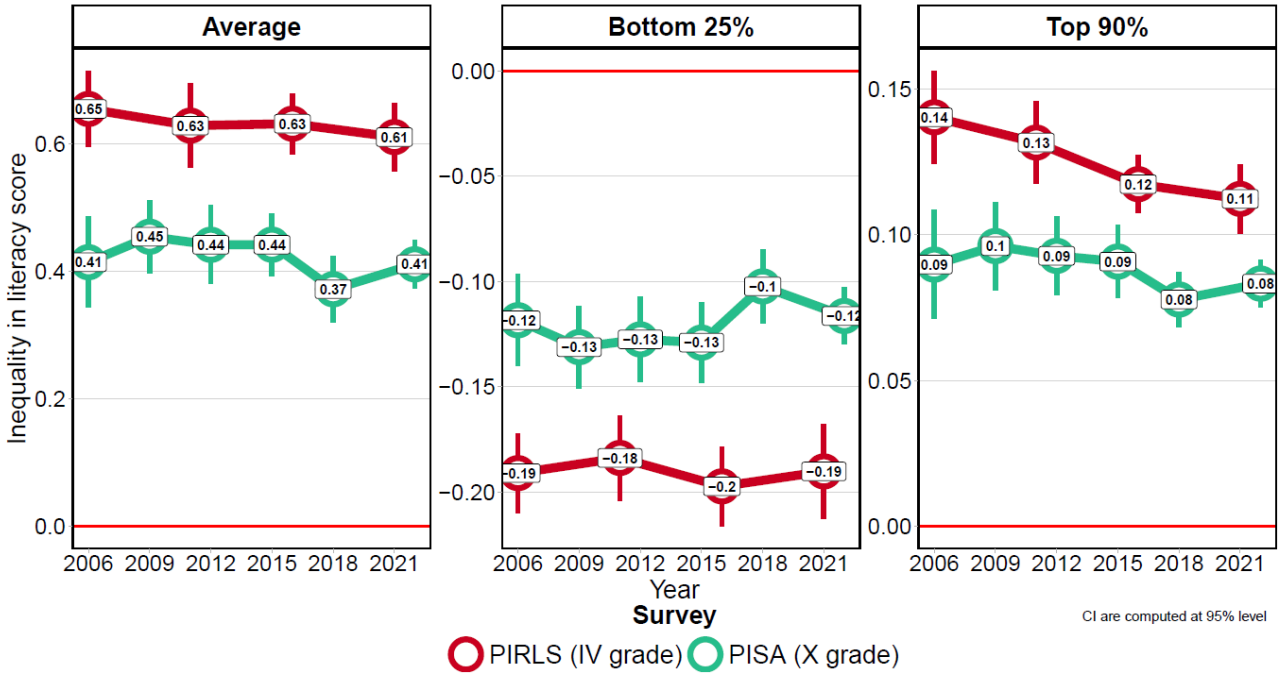
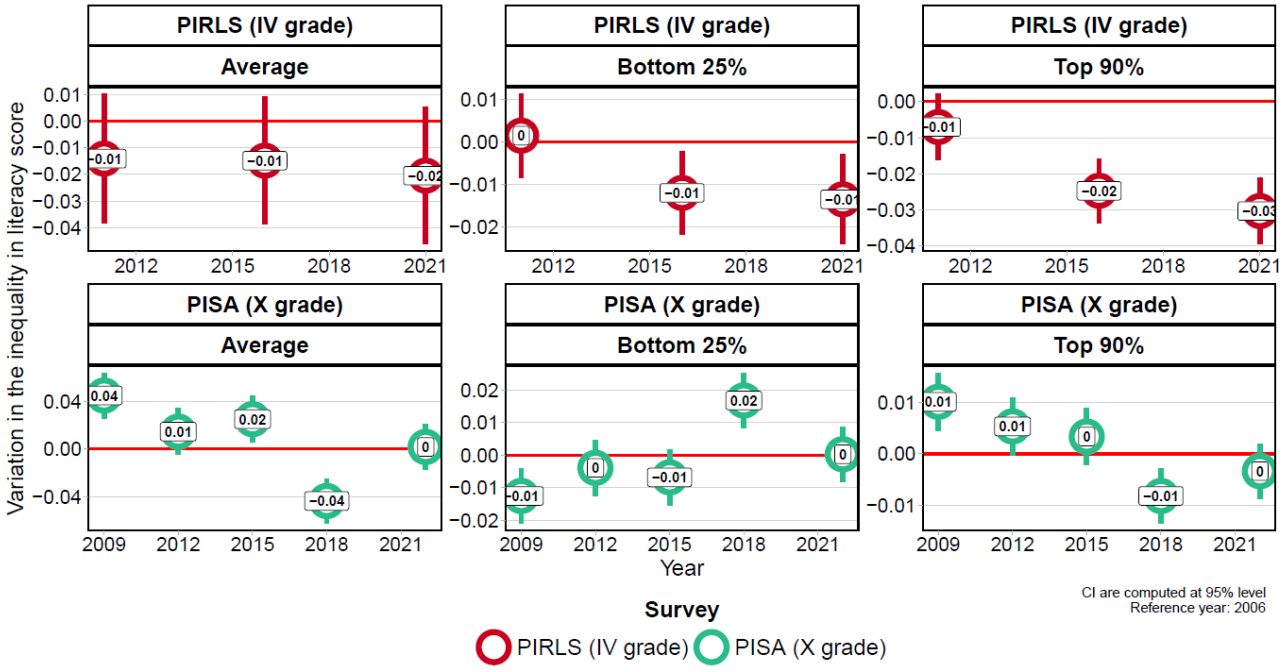


Figure 2. Yearly variation of inequalities in reading literacy achievement by survey. Reference year: 2006.



Keywords: Education, inequality, reading literacy, longitudinal

Intercultural Competences in the International Civic and Citizenship Education Study (ICCS)

Agostino Portera

First implemented in 2009, with follow-up cycles in 2016 and 2022, the current ICCS cycle (2027) continues to play a central role in investigating how young people are prepared to undertake their roles as citizens in contexts of democracy and civic participation. In a global and interconnected world, it is essential to gather data on students' knowledge and understanding of concepts and issues related to civics and citizenship, as well as their beliefs, attitudes, and behaviors. Particularly relevant for policy interventions and practical applications in the educational field—especially in school education—are contextual data on the organization and content of civic and citizenship education in the curriculum, teacher qualifications and experiences, teaching practices, school environment and climate, and support from home and the community.

Given the increasing influence of neoliberalism, and the realities of multiethnic and multicultural societies (Bauman & Portera, 2021), this presentation will be based on the results of a seven-year longitudinal case study. During this study, the life stories, conflicts, crises, and problem-solving strategies of 23 young people of Italian origin with migration experiences in Germany were observed (Portera, 1995, 2022). The findings highlight elements related to the inclusion of the second generation of immigrants in the host country and suggest that an intercultural approach in civic and citizenship education can play a central role in fostering enrichment and growth for young people living in democratic multicultural societies—rather than increasing difficulties, social disorders, or mental health issues. Achieving intercultural competences (IC) is essential for practitioners to implement culturally responsive practices and assessments, helping teenagers develop stable, flexible intercultural identities. The proposal will reflect on how to incorporate more items related to IC in the 2027 ICCS cycle.

Keywords: Intercultural Competences, Civic Education, Citizenship Education, International Civic and Citizenship Education Study (ICCS)

Equalizing Early? The Effects of ECEC Expansion on Achievement Levels and Gaps Across Countries

Giuliana Parente – Guido Salza – Moris Triventi – Nathalie Vigna

The expansion of early childhood education and care (ECEC) has become a cornerstone of the social investment paradigm, which promotes early education as a way to enhance human capital, foster social inclusion, and prevent cumulative disadvantage (Morel et al., 2012; Hemerijck, 2017). Rooted in developmental theories such as the "learn begets learn" hypothesis (Heckman, 2006), this approach holds that high-quality early learning experiences can lay the groundwork for later cognitive and socio-emotional development, with long-term consequences for educational trajectories.

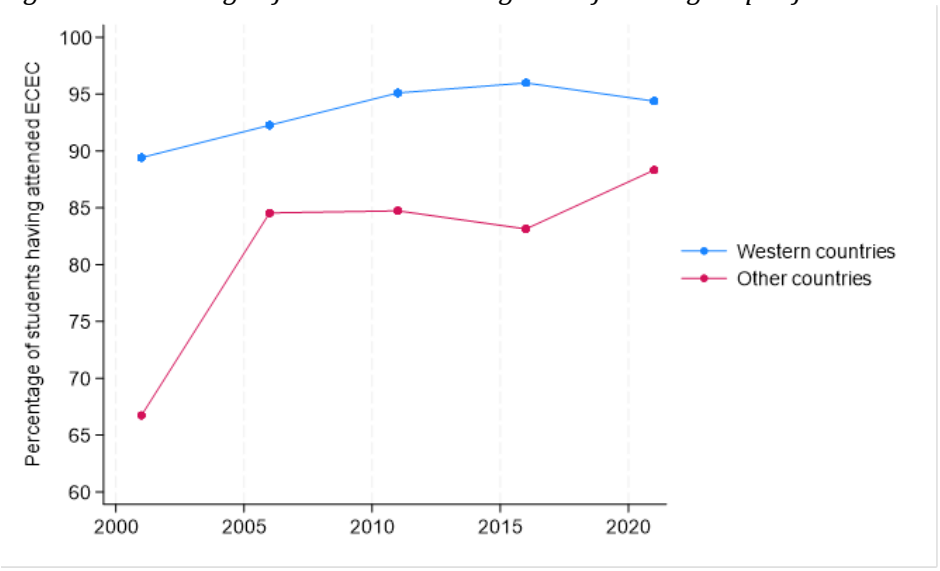
According to the counterfactual model of learning proposed by Raudenbush and Eschmann (2015), educational interventions like ECEC are expected to have stronger benefits for children from disadvantaged socioeconomic backgrounds, who typically lack cognitively stimulating environments outside formal education. By offering structured learning, rich language exposure, and emotional support, ECEC can potentially compensate for early developmental gaps and promote more equitable starting points at school entry. Furthermore, early education experiences, in the long run, can improve overall achievement and reduce social inequalities.

In recent decades, many countries have expanded access to ECEC to achieve both educational and social equity goals (OECD, 2020). Despite case studies in countries such as France, Germany, and Norway suggesting that ECEC expansion showed a positive influence on achievement and in reducing early disadvantages (Havnes & Mogstad, 2011; Felfe & Lalive, 2018;), empirical evidence on the effects of ECEC expansion remains fragmented, especially in terms of cross-national comparisons and the assessment of differential impacts on average achievement and inequality (Kulic et al., 2019).

This study aims to fill this gap by leveraging repeated cross-sectional data from five waves of the Progress in International Reading Literacy Study (PIRLS, 2001–2021) to examine how the expansion of ECEC is

associated with three key educational outcomes at the country level: (1) average reading achievement, (2) the prevalence of educational poverty (defined as the proportion of students performing below a basic proficiency threshold), and (3) achievement inequalities by social origin and gender.
Our main hypothesis is that higher participation in ECEC is associated with increased average achievement and reduced educational poverty. We also expect ECEC expansion to be linked to lower social inequalities in learning outcomes, though effects on gender disparities may be less clear-cut.
We use repeated cross-sectional data from five waves of the Progress in International Reading Literacy Study (PIRLS), conducted in 2001, 2006, 2011, 2016, and 2021. PIRLS assesses reading comprehension among fourth-grade students across countries. Our key independent variable is the national share of students who attended ECEC for at least one year before starting primary school, as reported by parents in the PIRLS background questionnaire.
We employ multilevel models with correlated random effects (Bell & Jones, 2015) to separate between-country differences from within-country changes over time. This analytical strategy allows us to identify both cross-sectional associations between ECEC coverage and educational outcomes and the effects of temporal changes in ECEC participation within countries.
Preliminary results indicate that participation in early childhood education and care (ECEC) has increased in nearly all countries, especially those with low initial coverage rates. This suggests a global convergence trend. Many Western countries have maintained high participation rates (above 90%) throughout the entire period (Figure 1). However, some countries have experienced fluctuations or declines in recent years. For example, Germany's participation rate declined from 99% in 2011 to 84% in 2021, and France's rate declined from 100% in 2011 to 90% in 2021. Notable increases are observed in non-Western countries, such as Azerbaijan (increasing from 42% in 2006 to 73% in 2021), Iran (increasing from 41% in 2001 to 85% in 2021), and Indonesia (increasing from 64% in 2006 to 73% in 2011).

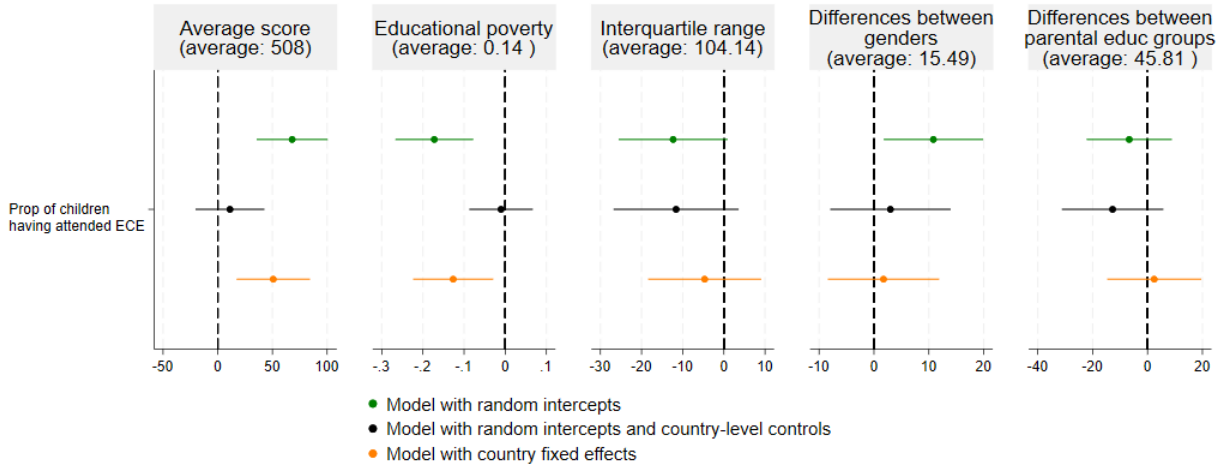
Figure 1. Percentage of children attending ECEC for two groups of countries



Note: the graph only includes countries observed in at least three of the five years

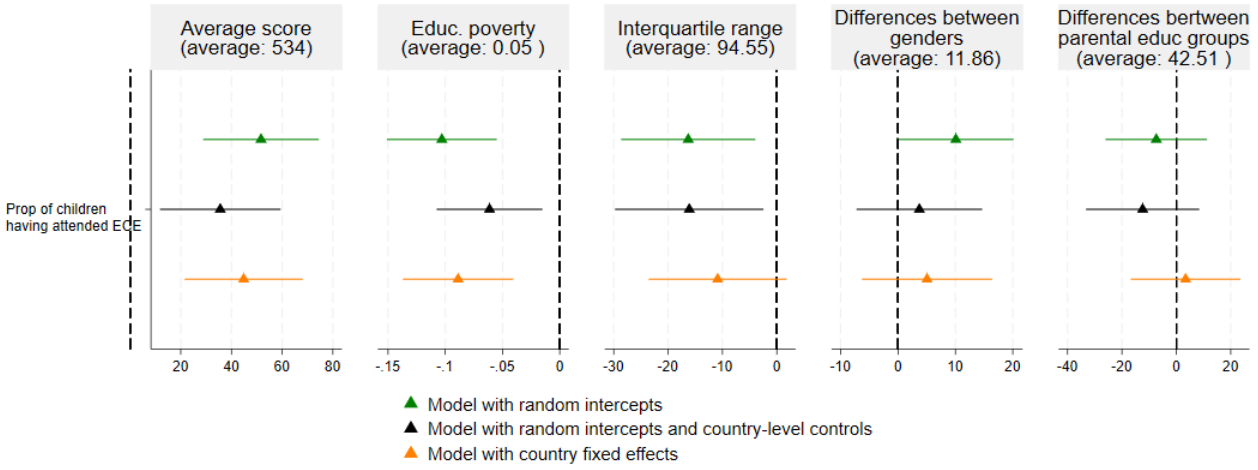
Figure 2 show the results from meta-multilevel models, which estimate the correlation between the proportion of children attending ECEC and several educational outcomes, including average reading achievement and indicators of inequality. Several models are estimated for each outcome, from the least restrictive (model with random intercepts) to the most restrictive (model with country fixed effects, relying only on within-country variance). The intermediate specification corresponds to a random intercept model including several country-level control variables such as the Human Development Index and the GINI index. We can see that ECEC expansion is significantly associated with higher average reading achievement indicating that countries with higher rates of ECEC participation tend to have better overall literacy scores among fourth graders. Furthermore, there is a negative correlation between ECEC attendance and educational poverty.

Figure 2. Results from the meta multilevel models: the correlation between the proportion of children having attended early education [in a specific country and year] and several indicators of achievement and achievement inequality.



These findings are consistent with those observed after restricting the sample to Western countries only (Figure 3): higher ECE participation is associated with **higher average achievement** and **lower inequalities**. These effects may be more pronounced in this subset, suggesting that in countries with stronger welfare systems or more established ECE infrastructures, the equalizing potential of early education may be more fully realized.

Figure 3. Results from the meta multilevel models: the correlation between the proportion of children having attended early education and several indicators of achievement and achievement inequality, only in Western countries



However, we do not find significant results regarding the relationship between ECEC expansion and inter-group inequalities. While ECEC expansion may boost overall student performance, it does not automatically reduce disparities based on students' social origin or gender.

These preliminary findings highlight the promise, but also the limitations, of ECEC as a policy tool for educational equity. The effects of ECEC likely depend on critical program characteristics such as quality, intensity, and the degree to which participation is targeted toward disadvantaged groups (Burger, 2010; Esping-Andersen et al., 2012).

In the next phase of the analysis, we plan to explore cross-level interactions between national-level ECEC coverage and individual-level attendance. This will help determine whether the broader institutional context moderates the relationship between individual ECEC participation and student outcomes.

Keywords: Early childhood education, educational inequalities, educational poverty, cross-national analysis

PISANU - DATA LITERACY AND DATA USE IN A DIGITAL AGE.
SESSION 12 - SUPPORT TEACHERS AND CO-TEACHING MODELS FOR INCLUSION: INVESTIGATING EQUITY IN THE ITALIAN EDUCATIONAL SYSTEM THROUGH CASE STUDIES AND COMPARATIVE PERSPECTIVES

ORGANIZER: UNIVERSITY OF BARI AND FBK-IRVAPP - INDIRE
COORDINATOR: SERAFINA PASTORE - DAVIDE AZZOLINI - SONIA MARZADRO - SARA MORI - SERENA GRECO - FRANCESCA STORAI
20TH NOVEMBER: 2.00 P.M. - 4.15 P.M. {AULA 4 – RESEARCH 10}

Understanding Data Literacy for Teaching: Insights from a Focused Literature Review on Practice in Schools

Sonia Marzadro – Davide Azzolini – Olga Gorodetskaya

Introduction

In the wake of accelerated digitalisation in education—especially following the COVID-19 pandemic—teachers are increasingly confronted with vast and diverse forms of data generated by digital platforms, learning analytics tools, and AI-enhanced educational technologies (Pangrazio & Sefton-Green, 2022; Michos et al., 2023). These data can inform pedagogical decisions, enable personalised learning, and support school improvement strategies (Mandinach & Gummer, 2016). However, research indicates that data use in schools remains fragmented and often isolated, and that teachers frequently lack the necessary skills, confidence, and support to make meaningful use of the data available to them (Brodsky & Brooks, 2020).

Data Literacy for Teaching (DLFT) has emerged as a critical competence for 21st-century educators and is increasingly recognised as a core component of digital competence by the European Commission (e.g. *Digital Education Action Plan 2021–2027*; DigComp 2.2, 2022). However, despite growing attention, DLFT remains poorly defined, unevenly implemented, and under-researched across EU education systems.

Research Aim

This paper presents the findings of a focused literature review conducted within the Erasmus+ project *EVIDALI – Evidence for Data Literacy* aimed at enhancing both policy and practice in DLFT. The aim of the review is to map and synthesise current knowledge on DLFT in school practice, with the specific objective of informing the design of a large-scale targeting primary and secondary school teachers. The review explores how DLFT is conceptualised, how it is operationalised in everyday teaching contexts, and how it can be effectively measured, with a particular emphasis on the real-world challenges and barriers teachers face in data use.

Research Method and Data Analysis

A structured literature review was carried out, analysing both peer-reviewed academic publications and grey literature on data use in schools. The analysis is organised into three interconnected analytical dimensions:

1. **Conceptualisation**: The review traces how DLFT has evolved from traditional practices—such as analysing test results or observing classroom behaviour—to more complex forms of data use enabled by digital and AI tools. Data in educational settings now include real-time process data, engagement metrics, and learning analytics, captured through platforms, apps, and online assessments (Jarke & Breiter, 2019; Michos et al., 2023). However, there is no unified definition of DLFT, nor consensus on what data should be prioritised or how it should be interpreted and acted upon by teachers (Mougiakou et al., 2023). Definitions of DLFT vary, and the review identifies a need for a multidimensional model encompassing diverse data types (static/dynamic; input/process/outcome), sources (administrative, digital, AI-generated), and purposes (descriptive, diagnostic, predictive, prescriptive) (Mougiakou et al., 2023; Mandinach & Gummer, 2016). Digital data enables unprecedented levels of insight into learners' behaviours and progress (Michos et al., 2023; Jarke & Breiter, 2019). Research highlights the importance of equipping teachers with data-related skills for improving instruction and student outcomes (Cavalluzzo et al., 2014; Comi et al., 2017), while also pointing out the lack of shared definitions and frameworks.

2. **Operationalisation**: The second section synthesises empirical studies exploring how DLFT is integrated into teaching. Key enabling factors include perceived usefulness, self-efficacy, and digital competence. The review draws on theoretical models such as the Theory of Planned Behaviour (Sadaf & Gezer, 2020), the Technology Acceptance Model (Scherer et al., 2019), and the Will-Skill-Tool model (Christensen & Knezek, 2001), which help explain variance in teachers' pedagogical uptake of digital technologies. Barriers include lack of training, low confidence, unclear objectives, and limited school-level support (Brodsky & Brooks, 2020). The literature also emphasises the role of socio-demographic and institutional variables—such as age, gender, digital skills, leadership, and professional development—in shaping how teachers use data (Michos et al., 2023).
3. **Measurement**: Finally, the review assesses available tools for measuring DLFT in teachers. While several survey instruments exist, most are limited in scope or insufficiently adapted to the complexity of AI- and platform-based data systems. The review identifies a set of validated instruments used in large-scale studies and extracts criteria for selecting or adapting tools (Scherer et al., 2019; Mandinach & Gummer, 2016). Emphasis is placed on instruments that can capture not only data-related skills, but also attitudes, values, and contextual enablers of data use.

Results

The literature reveals significant fragmentation in both the theory and practice of DLFT. While many teachers are expected to use data to improve teaching and learning, the conceptual and practical support they receive remains inconsistent. A critical gap persists between the growing expectations around data use and the actual training, resources, and tools available to teachers. The review proposes a structured framework to guide future empirical investigation and professional development, focusing on three core areas:

- A multidimensional model of DLFT that reflects the complexity of today's data ecosystems;
- Context-aware strategies for fostering meaningful data use in teaching;
- Robust, scalable tools to assess DLFT competencies and needs in diverse school settings.

These findings will inform the design and implementation of a survey on DLFT, which aims to generate large-scale, representative data on teachers' data practices, competencies, and training needs. Ultimately, the goal is to enable more coherent and evidence-informed support for DLFT across European schools.

Keywords: Data literacy, Data Literacy for Teaching (DLFT), educational data, digital education, teacher competencies, teacher training, evidence-based teaching

Data Literacy in the Evaluation Process of state-recognized private schools

Sara Pastore

Introduction

With the enactment of decree 80/2013 and the establishment of the National Evaluation System (SNV), since the 2014/2015 school year, schools have been required to compile a Self-Evaluation Report (RAV). The digitalization and systematization of this self-evaluation tool on a digital platform mark a turning point on two levels: on the one hand, it promotes the self-evaluation and, more broadly, evaluation processes, providing schools with essential elements for improvement; on the other hand, it assigns Digital Data a significant role, requiring teachers to acquire skills in collecting, managing, inputting, and interpreting data - in other words, to be trained in aspects of Data Literacy.

Data Literacy is a complex competence (Mandinach & Gummer, 2016) that integrates knowledge, the ability to use data, attitudes towards its use, the application of data based on goals, and the actions schools can take following their interpretation (Lee et al., 2024). These components, which define the concept of Data Literacy, are essential for compiling the RAV. INVALSI's reference framework enables schools to reflect on strengths and weaknesses and to collect useful data for planning improvement actions.

However, for the RAV and other strategic tools (Improvement Plan, Three-Year Educational Offer Plan, and Social Reporting) to become valuable resources for schools, school personnel must not only possess the technical skills required by the digital system but also recognize the potential of data. In fact, reading and analyzing data does not necessarily mean understanding its value and purpose (Smith & Doe, 2023), a condition that may reduce the RAV to a mere compliance exercise.

These considerations are particularly relevant during the transition from the 2022–2025 three-year period to the new 2025–2028 phase of the National Evaluation System, which includes the integration of the standard RAV with the Early Childhood RAV and, for the first time, involves the preschool segment. While staff in public schools may benefit from experience gained during previous RAV compilations, many staff members in private schools — including teachers and coordinators of educational and teaching activities - will face the data collection and self-evaluation process for the first time.

The upcoming 2025–2028 scenario introduces a significant change for preschool institutions and, more broadly, for private schools that, although already involved in the 2022 census (DGOSVI No. 13483 of May 24, 2022), have not yet participated in the evaluation process as per Article 6 of Decree 80/2013.

Additionally, the DGOSVI note No. 21153 of May 29, 2025, titled *"SNV - Census of Private Schools for the Preparation of Strategic Documents,"* reiterates the importance of these schools registering on the digital platform, which is essential for compiling strategic documents. The decision to investigate the training needs of private schools stems not only from recent regulatory changes but also from observations of data from the previous three-year period (2022–2025): only 34% of private schools published their RAV during that cycle, compared to 97% of public schools (data provided by the Ministry of Education and Merit).

Given this historical and theoretical background, a research study was launched to support schools — particularly private institutions (micro level), Regional Education Offices (meso level), and the education system as a whole (macro level) — by fostering not only technical skills for initiating the self-evaluation process but above all, the development of a data culture. According to the proposed model, training designed on this basis could have direct/indirect and short/long-term effects:

1. Short-term direct effects: operational capability in consciously compiling the RAV;
2. Long-term direct effects: sustained ability to consistently compile the RAV;
3. Short-term indirect effects: increased motivation and perception of self-evaluation's effectiveness during compilation;
4. Long-term indirect effects: long-term motivation and perception of effectiveness, development of a Data Culture (Hilty et al., 2025) and Evaluation Culture (Mayne, 2009) as valuable support for organizational and planning thinking.

Research Methodology and Tools

Based on the hypotheses introduced, the research followed a QUAN–QUAL mixed-method approach. The first phase involved a quantitative analysis of 275 cases (teachers, coordinators, administrative staff, and other personnel) from state-recognized private schools who participated in the INVALSI investigation during the seminars titled *"The RAV and the National Evaluation System: Guidelines for the 2022–2025 Triennium."*

This was followed by a qualitative exploratory phase, including a semi-structured interview with the Ministry of Education and Merit (MIM), as a key stakeholder, and three focus groups conducted in the 2024/25 school year in three privately managed primary schools, during the pre-test phase of the integrated RAV by INVALSI. The interview with the MIM aimed to explore the Ministry's perspective on future policies regarding strategic tools, feedback received from private schools on the census and RAV compilation, and training needs for the upcoming 2025–2028 period.

The three focus groups concentrated on schools' previous experiences with self-evaluation and their training needs, to understand the specificities of state-recognized private schools regarding the process of self-evaluation and strategic document compilation. Each group interview lasted about two hours, was led by two INVALSI researchers (moderator and observer), and involved the Internal Evaluation Committee (NIV) and the Coordinators of educational and teaching activities from each school.

In the three schools involved, the NIV consisted of 3 or 4 teachers, maintaining the structure of the previous triennium, with the replacement of those no longer in service. As emerged during the focus groups, private schools are often seen by teachers as transitional workplaces while awaiting placement in public schools.

Potential biases were addressed during the interview design phase, which focused on two main areas: the self-evaluation process (where present) and training needs and expectations.

Results

The secondary analysis of the data collected during the seminar cycle *"The RAV and the National Evaluation System: Guidelines for the 2022–2025 Triennium"* was conducted on responses from 275 participants from private schools. The analysis revealed that 68% expressed training needs related to self-evaluation, the compilation of strategic documents (with particular focus on each document's specific characteristics and their interrelations), and the interpretation of data on the platform.

The survey module used in the seminars included an open-ended question that allowed qualitative analysis. In many cases, participants answered vaguely (e.g., *"a bit of everything…"*), suggesting that part of the school staff in private institutions may lack adequate preparation on self-evaluation and strategic documentation and may feel disoriented during the process.
This sense of disorientation was also reflected in the interview with the MIM representative, who noted that state-recognized private schools often struggle with the aggregation process required by the census. The interview revealed that many state-recognized private schools share the same Coordinator, and the platform has been designed to accommodate this by allowing school aggregation under a single institutional code and unified RAV compilation.
However, this aggregation process and the acquisition of a unique institutional code were unclear to many institutions, leading to wasted effort and multiple RAV compilations — an onerous task given the limited number of teachers in state-recognized private schools.
To identify specific training needs, the focus group findings were invaluable in highlighting both the content and timing (before, during, and after the RAV compilation) and the preferred modes of delivery (mainly digital).
The pre-test phase proved useful in promoting a more informed perspective on data and in providing, through the focus groups, critical insights on the self-evaluation process and the empowerment opportunities offered by Digital Data.

Conclusion

In line with the seminar's title, this study provides useful insights into how processes can transform data "of" the education system into opportunities "for" the education system, supported by tailored guidance based on the specific needs of schools.

Keywords: Data literacy, data culture, evaluation, teacher training

Large-scale Assessments' Data-use: How Support Teachers' Assessment and Data Literacy

Serafina Pastore

Since the last century, the empirical turn, stressing the desire for educational quality improvement, led to a quest for data and information about education. The interest reserved for student learning assessment, as well as the growing emphasis on standardized tests and high-stakes assessment, have induced national school systems to diversely frame assessment to reply both to accountability requirements and educational improvement (e.g., Australia, Canada, South Korea, Hong Kong) (Smith, 2016). Despite the traditional resistance and persistent confusion around their purposes and relevance (Stiggins, 2017), large-scale assessments (LSAs), at national and international level, have been progressively recognised by school administrators and school personnel as a component of educational governance and instructional quality systems (e.g., allocation of financial resources, commitment of policy makers to improve national curriculum; need for continuous teacher professional development).
Over the years, the 'broader, deeper and consequential' (Emler et al. 2019, 280) use of LSAs was justified by the need for information about student-based decisions and the evaluation of educational policy, national curricula, and teaching practice. However, while LSAs have been acknowledged as relevant components of educational accountability systems, teachers' negative attitudes towards LSA programmes and the lack of assessment literacy have been highlighted (Fullan et al., 2018; Klinger & Rogers, 2011). Different studies have pointed that teachers encounter difficulties in understanding the value of LSA (DeLuca & Bellara, 2013; Leighton et al. 2010; Pastore, 2023). More specifically, recent studies on the use of assessment data for decision-making and teaching practice have showed that although teachers recognise the importance of using evidence and data gathered through assessment, sometimes, they are not able to manage several sources of information including data from LSAs (Farrell & Marsh, 2016; Mandinach & Gummer, 2016; Schildkamp et al., 2014). Consequently, teachers do 'not necessarily ensure assessment literacy or the knowledge needed to engage in using assessment data to inform instruction' (Will et al., 2019, 1).

Although the effects of LSAs beg for a reflection on how actors use these data for national agenda setting on educational policy innovation, a critical revision of education initiatives purposed to support teacher assessment literacy, data-use, and data-based decision-making is required.
Assuming a complexity theory framework, the present paper focuses on professional development tools, practices, and initiatives that, over the last 20 years, aimed to orient teacher assessment practice toward the use of LSAs' data. Encompassing the theoretical and practical plans of educational research, policy, and practice, the present paper proposal intends to review and scrutinize under which conditions, why, and how LSAs, teacher data literacy, professional development and assessment practice interact. This contribution, framing the theoretical framework of the session proposal "Data Literacy and Data Use in a Digital Age" arguments in favour of investigating the use of LSAs by considering practical assessment challenges for teachers and their assessment conceptions which are of paramount importance to ensure teacher assessment literacy, teacher professional development, and effective school improvement.

Keywords: Large-scale assessment, teacher assessment literacy, data literacy, professional development

Evaluation, Recognition, and Improvement: Using Student Questionnaires to Enhance Teaching Quality in Upper Secondary Schools

Francesco Pisanu – Paolo Pancheri – Giuseppe Rizza – Francesco Rubino

In recent years, the Autonomous Province of Trento has implemented a structured use of data collected through the "student customer questionnaire" (Provincial Evaluation Committee, 2024), administered annually to approximately 5,000 upper secondary school students. Originally designed to assess general satisfaction, the questionnaire includes a section specifically focused on teaching evaluation, with items inspired by the framework developed by the Bill & Melinda Gates Foundation within the Measures of Effective Teaching (MET) project (Kane & Staiger, 2012). This internationally validated model aims to capture key dimensions of teaching quality from the student's perspective, such as clarity of explanation, engagement, individual support, and respect.
The results of these questionnaires are annually returned to individual schools in aggregated form, serving as a valuable resource for teacher team reflection, the drafting of the school self-evaluation report (RAV), and the definition of improvement priorities. In some schools, the data has been reinterpreted and embedded in broader decision-making processes, giving rise to pedagogically informed data use practices.
This contribution focuses on two levels of analysis. The first explores provincial-level data, collected annually since 2018 from about 5,000 upper secondary students. The quantitative analysis examines the distribution of student ratings on teaching quality, with particular attention to internal consistency and variation between schools. Recurrent patterns emerge in students' perceptions of teaching, highlighting both strengths and areas for improvement, and reflecting the alignment between perceived teaching and student needs.
The second level presents a case study of a school that, since 2018, has implemented a system for providing individual feedback to teachers, using the data to determine the merit-based bonus, as outlined in provincial regulations. Here, the data (mainly derived from a set of items similar to the provincial questionnaire but teacher-specific) are complemented by qualitative documentation produced by the school to support and sustain this practice. The disaggregated data are returned and discussed within the teachers' assembly, fostering transparency and legitimacy of the process. In this context, teachers are not only evaluated but also actively engaged as recipients and interpreters of the data in a professional development framework.
The analysis shows how data use culture in schools can evolve toward a formative approach, promoting reflective and responsible teaching practices (Argentin, 2018). When properly interpreted and shared, provincial questionnaire data foster widespread data literacy among teachers and inform instructional improvement (Datnow & Hubbard, 2015). The case study further demonstrates that the use of data for performance-related purposes can be balanced when supported by educational leadership, validated tools, and participatory processes. The findings suggest that combining quantitative data with qualitative sources is essential to trigger robust and sustainable evaluation processes that balance accountability and professional development. In this perspective, the school leader plays a key role in mediating between data, interpretation, and decision-making (Paletta, 2014).

This contribution aims to stimulate discussion on the role of data in promoting teaching quality, offering a concrete example of integrated, multi-level use of student evaluation data with relevant implications for research and education policy.

Keywords: Teaching evaluation, Student questionnaires, Teacher data literacy, School improvement

Mathematics as tool for artificial intelligence literacy

Maria Cristina Carrisi – Sara Vergallo – Mirko Marras

Younger generation is growing up in a world increasingly shaped by technologies and artificial intelligence (AI), and it is now the most common tool through which individuals interact with technology and process information. Which makes literacy in these tools a critical resource for developing the skills needed to understand and use them critically. Without a deep understanding of AI, new generations risk passively interacting with technologies, carrying misconceptions, unrealistic expectations, and an inability to evaluate the algorithmic results they receive as output (Randi, 2018, Yang, 2022). AI education should therefore start in school but, in current textbooks and school practices, the importance of data (and its use) is underestimated. Instead, we believe that understanding how to read, use, and process data is a key element in understanding AI-based systems (Olari, 2020). We also believe it is fundamental that, a path to AI literacy, must necessarily be preceded by the consolidation of certain mathematical topics including: classification, data representation, the use of fractions and their comparison. Although it is widely documented in the literature how technologies play a crucial role in shaping mathematical thinking and knowledge production (Borba & Villareal, 2005), there is, however, a gap in the research on the opposite mechanism: how to improve the learning of computer science through an enhancement of related mathematical content?

The draft of the new national directions (MIM, 2025), for the first time, states: "Mathematics is also a cross-curricular discipline. It is present in all sciences and is the basis for computer science, a fundamental subject for consciously facing an increasingly digital and increasingly less physical world". In this unofficial document, computer science is introduced, for the first time, both as its own subject and as part of mathematics, emphasizing the intimate connection between the two disciplines.

Thus, in this paper we propose a path for AI literacy intended for primary (last year) and secondary (first year) school students, based on strengthening mathematical skills closely related to the use of data and, in any case, present in elementary school curricula. Our aim is to strengthen AI conceptualization, data representation, classification and evaluation of AI-processed results. To evaluate the effectiveness of our approach, we have so far conducted two experiments on a total of three fifth-grade elementary school classes and two first-grade secondary school classes. During the activities we have evaluated their progress, as well as in itinere through critical observation and a logbook, by comparing a pre-test and a post-test and administering a final satisfaction survey. The overall goal of our work is to demonstrate how primary education can effectively integrate fundamental AI concepts with mathematics to improve students' understanding and active engagement. We investigate how a structured learning pathway can link AI and mathematical reasoning by emphasizing the importance of topics such as classification, set theory, and data representation. Therefore, we designed a hands-on curriculum, which can be offered at different periods of the educational schedule, also introduces the theoretical and mathematical foundations of AI through interactive and problem-solving activities. The questions we asked ourselves before we began were: do children have enough foundation (mathematical and cultural) to tackle a simple computer science curriculum? How can this knowledge be improved and, at the same time, improve children's sentiment in stem subjects? What improvements can I achieve, how and how to measure them? Our research currently focuses on evaluating the impact of this approach on students' understanding of AI, its role in strengthening math skills, and its effectiveness in promoting engagement and interest. To analyze these aspects, we moved on three levels: first, we presented a new learning pathway that systematically integrates AI principles with mathematical concepts, ensuring alignment with elementary school curricula. Second, we described and explained in detail the implementation of this learning pathway, including the selection and preparation of teaching materials, which were carefully curated to support students' cognitive development in both AI and mathematics. Third, we have evaluated (and continue to do so) the effectiveness of the learning path

through a study involving several class groups so far, analyzing both quantitative and qualitative results to assess the improvements and engagement achieved. We wanted to monitor:

- The improvements achieved in comparing fractions and different methods of representing data (Euler-Venn, double-entry tables, tree diagrams)
- The changes between initial and final opinion on computer science, AI, and mathematics.

Our approach is designed to provide a coherent and progressive learning experience. The course is structured to reinforce prior knowledge, guide students in identifying AI limitations, and introduce new models of representation to support cognitive development. The underlying theories are those of constructivism (Jonassen, 1994) and constructionism (Papert, 1971), and we adopted a spiral learning approach (Bruner, 1960), ensuring that concepts were reintroduced at increasing levels of complexity. To implement this path, we combined original teaching materials with resources adapted from established educational settings. The learning modes are structured following the learning-by-doing and learning-by-necessity methodologies. These strategies encourage students to actively experiment, refine their knowledge, and engage in iterative problem solving, with targeted interventions on our part whenever previous approaches have proven insufficient. In addition, the use of different semiotic representations (Duval, 2006) facilitates students' understanding of classification concepts through multiple representational models, e.g., Euler-Venn diagrams, tabular data structures, and decision trees. Topics and themes are chosen based on four of the "Five Big Ideas" (Touretzky, 2023), namely perception, representation and reasoning, learning and social impact.

To design the activity, we relied on ministerial directions and, in particular, the learning objectives of the area of “relations, data and predictions” at the end of fifth grade in elementary school. Regarding pre and post tests, we found it useful to include two questions taken from INVALSI tests of past years, selected from the gestinv archive, in order to have an objective and validated reference point. These are the following items:

- Mat- SNV 2012 05 02-B
- Mat- DR 2018 10 01-A

Both pertain to the representation of data. We selected these items because they require the ability to use certain knowledge and skills in a different and differently articulated context and because they ask students to move between different semiotic representations (Duval, 2006). Although the second item chosen is designed for a grade 10, we felt it required a level of proficiency (01) low enough to be administered in a fifth grade primary.

From the analysis of the papers of 21 primary students and the observation in itinere, we can briefly state that:

- Students demonstrated a solid general understanding of artificial intelligence concepts: many subjects obtained high scores during the proposed exercises.
- Students showed a satisfactory foundation in mathematical reasoning, particularly in interpreting frequencies and solving problems through sets.
- The activities were overall successful in promoting engagement and interest. While students generally found the learning path fun and educational, they refined some activities to ensure sustained engagement.
- Students stated that, such activities, helped them understand the usefulness of mathematics in different contexts.

Analysis of the responses provided by primary students specifically to the INVALSI questions we administered revealed some interesting evidence:

- Mathematics was found to be a useful tool for conveying information regarding how machine learning works; however, improvements in the tests were more on the computer aspect than the mathematical aspect
- Fifth-grade students have some critical issues in the representation of data using Euler-Venn diagram: in the first administration of the item, no child was able to represent elements outside the pre-prepared sets, thus in the universe set. In the second administration, after working on classification and data representation aimed at understanding the functioning of the ML, 10 percent of the children correctly placed all elements within the proposed representation.
- The grade 10 item was performed correctly by 33% of the children in the pre-test and 66% in the post-test, after working together on analysis and construction of dichotomous trees.
- As for the grade 5 item, 14% answered all 4 questions correctly in the pre-test, compared with 78% in the national results. In the post-test, there was no significant improvement. The question with the most correct

answers was 3 in both tests and the question with the fewest correct answers was 1, in both, probably due to problems with understanding the question.

Further analysis involved comparing the improvements made in the primary with those made in the secondary.
The conclusions drawn from the data analysis were very valuable for rethinking the design of the intervention and the aspects on which to dwell most during the activities, namely: correctly representing the absence of one or more features in the Venn diagram, calculating the accuracy of a machine learning model, and finally recognizing, in a group of items to be classified, the determining features for the creation of dichotomous trees.

Keywords: AI literacy, classification, data representation, math education

Towards an ecology of inclusion

Graziella Arazzi – Marina Orselli – Paola Giannoni

Introduction
The driving force behind the survey was the implementation in Liguria, in the two-year period 2023/2025, of the INDIRE Project - Small Schools "Classes on the Net" which - as is known - through the matrix of technological innovation, intends to reduce educational gaps and multiply the opportunities of school communities (multi-classes, isolated complexes) located in marginal territories (inland areas, hinterland). The path, which involved 15 comprehensive institutes in the region (8 schools in the province of Imperia, 3 in the province of Genoa, 3 in the province of Savona, 1 in the province of La Spezia, with the participation of primary and secondary school classes and the participation of teachers of the disciplines covered by the national standardized surveys), it has become an engine for rethinking the quality of inclusion, representing a lever to redefine co-teaching and the relationship between homologous but still different professional figures. In the first phase, the training/experimentation/didactic innovation path was aimed at teachers on common posts. In a second period (beginning of the 2024/25 academic year), a series of managerial but also communicative difficulties, encountered by the working groups both in the practice of twinning/alliance/pairing at a distance of schools belonging to different territories and in the planning of activities with pupils, led school principals and project representatives to involve support teachers (10 teachers on a regional scale). With a view to supporting the implementation of the project, the audience of 25 school operators (distributed in 10 working groups) developed a shared scenario of analysis and discussion of some problems:
a) while generally stating the effectiveness of inclusion processes and self-evaluating themselves in the RAV at medium-high level for section *3.3.a Inclusion activities* (Processes – educational and teaching practices), the schools involved in "Classes on the Net" presented in a clear and ascertainable way the determining features of quality inclusion, highlighting the sustainability of the strategies used (Canevaro, 2019)?
b) What was the representation of the support teacher in micro-educational realities?
c) How was it possible to foresee the co-evolution of the different types of teachers, within a new design of co-teaching and in view of an unprecedented perspective of inclusive school organization, promoted by "Classes on the Net"?
A workshop of comparison and discussion at a distance between different professionals (referents and teachers of the twinned groups; support teachers) found a common theme in one of the research questions of the "Classes on the Net" project: *Q1 - Does the Classes on the Net model promote educational inclusion processes in terms of opportunities for small schools to overcome social and cultural isolation?*
With the aim of overcoming the stereotypes of technicality in the implementation of inclusion and the trivialization of professional interaction between teachers of various types, the working groups have identified two lines of research:

1) reflect on the various declinations of co-teaching, in situations of territorial decentralization and within non-standard schools;
2) explore the sustainability of school inclusion as a set of creativity, transformative power, potential to face risks, uncertainties, problems.

Subject, objectives and research hypotheses

The survey, developed by the 15 schools, in collaboration with the Regional School Office for Liguria and the Comprehensive Institute of Genoa Pegli, a school that is the management coordination hub of "Classes on the Net" in Liguria, defined the following object: factors and approaches for the development of proximity, ecological and sustainable support (widespread inclusion), with particular reference to the co-teaching system, in the three dimensions that characterize it within the pedagogical culture of the French-speaking area (Québec), from which the "Classes on the Net" path is borrowed:

1) Co-teaching in unitary space:
- tandem teaching;
- one teacher teaches, the other helps high-level groups;
- both work on small nuclei.

2) Co-presence (one teaches, the other observes)
3) Co-interventions in differentiated spaces:
- teachings in/of *ateliers*;
- teaching in differentiated groups;
- parallel teaching.

Objectives

Macro-lens:

To point out how the implementation of "Classes on the Net", integrated with innovative forms of self-evaluation of schools (participatory self-evaluation within and in the territory; self-evaluation conducted among several schools and with the contribution of families, associations, local authorities), can determine the overcoming of the homogeneous and abstract model with which the school institution, in the face of the problems of pupils in difficulty, believes that it can solve them independently, not considering some priority elements for inclusive processes: - interaction with multiple social worlds; - network decision-making processes; - shared representation of obstacles, solution ideas, failures; - reorganization of governance (collaboration between the instrumental function for inclusion, members of the Internal Evaluation Nucleus/NIV, representatives of "Classes on the Net").

Sub-objectives:

highlight how inclusive practices translate into mere bureaucratic fulfillment if inclusion is not linked to the following process objectives: 1) development of a care context that takes care of fragility and vulnerability not only of learners but of all the social actors involved (teachers, administrative staff, parents); 2) testing of designs and models for the evaluation of inclusion in contexts characterized by unprecedented situations, unpredictable phenomena, perturbations.

Identify new training contents - currently mostly disregarded (Baggiani, 2024) - for the small school presidium of inclusive culture, with attention to: - different prototypes of teaching professionalism, in the contextualized ("tailored") application of the European standards of the teaching profession (DM 226/2022, art. 4 - Training of newly hired teachers); - integration between digital culture / *Embodied Education* (already consolidated) /*Embodied Formation* (in its early days).

Research hypotheses

"Classes on the Net" - in the multiplicity of implementation plans and in the variety of outcomes - has constituted a sort of experimentation-accompaniment in itinere for:

- ✓ redesign the function of support teachers and at the same time enhance the other figures (teachers on common posts, school managers);
- ✓ orient the school community to differentiate strategic choices, translating some elements of *the Manifesto for Inclusion* (2017) into management and organizational paths, including: "The construct of inclusion requires multidimensional analysis and interventions and the concurrence of multidisciplinary, interdisciplinary and transdisciplinary visions and readings" (point 3 – *Inclusive contexts*).

Data used

With a view to a comparison between the 15 comprehensive institutes, with regard to the reading of the results of primary and lower secondary school students (attending in the complexes involved in "Classes on the Net") in the INVALSI standardized tests, the following data were used, present in the "Outcomes" section of the RAV:

- ✓ Test score and differences from schools with similar ESCS (2.2.a Student results in standardized tests).

- ✓ Variability of scores between classes and within classes (2.2.b Variability of results in national standardized tests).
- ✓ Effect of school on student results in tests (2.2.c School effect).

In view of a self-assessment of the schools developed in a participatory mode, the RAVs of the 15 schools were investigated, with a focus on the "Inclusion and differentiation" area (Processes – Educational and didactic practices) and examination of the indicators that mark the development of inclusive paths (data acquired through the "School Questionnaire").
In addition to these data, the following dimensions were collected and interpreted:

- n. and contents of the redesigns produced in the groups of the "Classes on the Net" project;
- n. and type of remote interactions between teachers participating in the project.

Method or approach

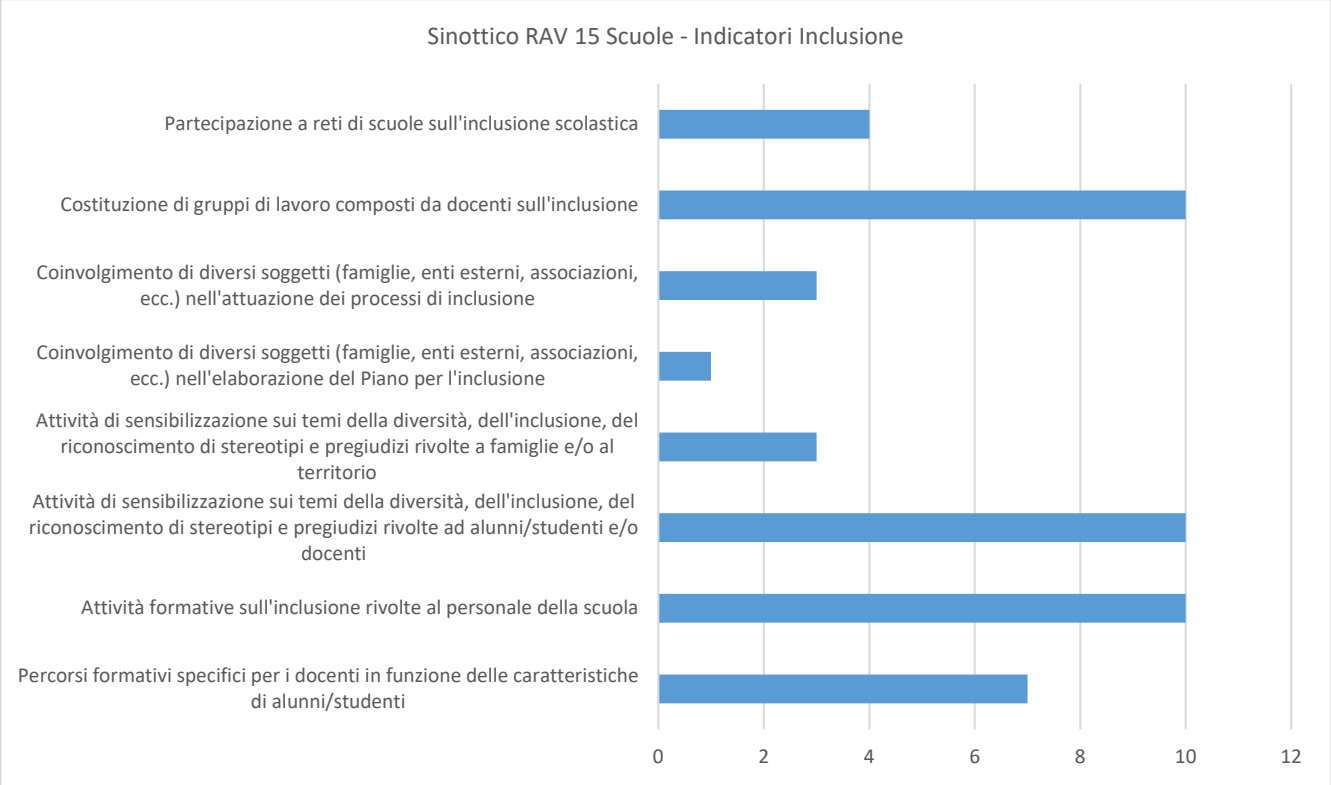
The qualitative approach entailed:

a) comparative documentary analysis of RAV and PTOF, published on the https://unica.istruzione.gov.it/portale/it/scuola-in-chiaro platform;
b) activation of remote exchanges on the explored topic. The interactions involved, in distinct phases: - some groups of "Classes on the Net"; - all 10 Groups of the project;
c) online interviews, by USR Liguria and with identical structure, with support teachers and teachers on common posts. The interviews were carried out and documented by the various groups of "Classes on the Net";
d) accrual budgets;
e) visiting activities.

Results or argumentation

Where the school - during "Classes on the Net" - has developed differentiated forms of interaction with territories, families and associations (awareness-raising activities for different social actors; stipulation of agreements; adoption of Plans/Protocols for inclusion), in addition to the indicators present in the RAV to intercept the quality of inclusion, the participants in the research are able to report new indicators, oriented to connote an inclusive system that expands in multiple scenarios and times, allowing a glimpse of the coevolution of the various professional figures and organizational flexibility. From the documentation of exchanges and interactions, it emerges that the teachers – struggling with management difficulties, obstacles to the project – have often recovered, through twinning with teachers from other complexes, internal multimodal spaces for dialogue/argumentation/decision. In parallel with this process, school principals, in some cases, have changed the composition of the working groups and strengthened the staff figures of the comprehensive institute.
A shared verification of the quality of inclusion within the 15 schools was also implemented, leading to the following dimensions, transversal to the teaching professionalism both on common post and on support: - exchanges of knowledge; - construction of guidance practices through the IEP (Guerini & Sannipoli, 2024); - opportunities and models of professional reflexivity; - management of "an ecosystem of supports" (Canevaro, 2019).
On the part of all schools, it was noted that in the self-evaluation of a. s. 2023/24 co-teaching was not investigated in all its articulations. Following the collaborative investigation, however, there is greater awareness in the analysis of the various prototypes, questioned in their practicability.
To respond to management needs (training modules for pupils not participating in the project), the complexes have implemented micro-experiments of co-intervention in parallel and in different places. All the participants in the research do not always consider co-presence effective with regard to multi-classes, unlike what is declared for standard classes. Called upon to pronounce on co-teaching, in accordance with the scientific literature on the phenomenon (Tremblay, 2024), teachers and principals express the following vision:

- Differing from co-intervention, co-teaching also seems to have an impact on the professional development of school figures;
- however, the cost/benefit ratio limits its practicability;
- territorial alliances appear necessary, guarantors of sustainable inclusion.



Keywords: Co-teaching, teaching professionalism, subsidiarity, enation

Support teachers and the educational community: a systemic interpretation of professional identity and training development

Sara Mori – Francesca Storai – Serena Greco

Introduction

In recent decades, school inclusion has established itself as a guiding principle of European education systems, receiving widespread international recognition, thanks in part to the Italian model, which is considered a positive example (Marchisio et al., 2024). Italian legislation, starting with Law 517/1977 and reinforced by subsequent decrees and guidelines (MIUR, 2009; Legislative Decree 66/2017, Guidelines for school inclusion, 2022), has promoted tools and guidelines to ensure equal opportunities and the full development of the potential of people with disabilities. However, the affirmation of the principle of inclusion is not in itself a guarantee of quality and equity in educational pathways (Cottini, 2016; Dell'Anna et al, 2023), particularly for the most vulnerable pupils. In this perspective, the role of the support teacher takes on crucial importance, not only as a figure supporting students with disabilities, but as an agent of equity and co-constructor of inclusive contexts within the entire school community. Yet, the professional identity of the specialised teacher is still ambiguous and sometimes marginal with respect to the various dimensions of school organisation. Fregola and De Mutis (2023), drawing on Pavone (2017), highlight how the term "support" is now mainly associated with the teacher in charge, rather than with a structured set of educational and planning actions in favour of inclusion, as required by law. This semantic shift risks confining teachers specialising in disabilities to a peripheral role, rather than valuing them as promoters of a truly equitable school system capable of responding to a variety of educational needs.

Starting from these premises, the aim of this paper is to investigate the perception that specialised teachers have of their role within the school, adopting a systemic perspective (Ruzzante, 2024). The survey is developed along three main dimensions: the educational dimension, with reference to co-teaching practices in the classroom and their effectiveness in promoting inclusion; the organisational dimension, which considers the recognition of support teachers as strategic figures in the school inclusion system; and the

territorial dimension, understood as the ability to activate and participate in collaborative networks with associations and institutions in the local context.

Main objectives:

1. To analyse the perception that trainee support teachers have of the possibility that their role can have an impact on implementing a shared plan with curriculum teachers within the classroom, so as to encourage inclusive pathways.

2. To analyse the perception that trainee support teachers have of the possibility that their role can contribute to the development of an inclusive school culture, from a systemic perspective.

3. To analyse the perception that trainee support teachers have of the possibility that their role can contribute to the development of the school's relations with the local community, with a view to creating a school that is increasingly connected to its environment.

Research hypothesis:

It is therefore hypothesised that trainee teachers recognise the significant potential of their role in the educational sphere, understood as the ability to initiate and support co-planning practices with curriculum teachers. If structured in an authentic way, this collaboration can be a lever for transforming the classroom into a truly inclusive learning space, where differences are not only accepted but become resources for the whole class.

With regard to the organisational dimension, it is assumed that trainee support teachers perceive their role not only as a specialist function aimed at pupils with disabilities, but as an integral part of an inclusive school system. It is expected that awareness will emerge of the strategic potential of their role in contributing to a collegial culture capable of promoting shared practices and a pedagogical vision centred on the educational rights of each and every individual.

In terms of the territorial dimension, it is assumed that teachers in training will see their role as a bridge between the school and the local community, recognising the importance of networks with families, services, associations and local institutions. In this perspective, support teachers can become facilitators of meaningful connections between the school and its context, helping to build a broader and more co-responsible educational environment.

Data used and analysis methodology

Data collection will be conducted through a questionnaire aimed at support teachers in training who are currently enrolled in extraordinary TFA courses. The questionnaire will consist of a combination of closed and open-ended questions in order to collect both quantitative data, useful for identifying general trends, and qualitative data, aimed at deepening teachers' perceptions, representations and meanings. The responses will be analysed in relation to the three dimensions investigated: teaching, organisational and territorial.

Based on these results, a SWOT analysis will be carried out for each of the three dimensions in order to identify strengths, weaknesses, opportunities and threats related to the perception that trainee teachers have of their role. SWOT analysis is a useful tool for integrating empirical evidence and critical reflection (Jonhson, et al, 2017), supporting the development of practices consistent with improvement objectives, in this case the trajectories of training courses. It will thus be possible to outline targeted strategies for the implementation of these courses, with particular attention to the development of professional skills consistent with a systemic approach to school inclusion.

Results and conclusions.

Data are currently being collected. In light of the three dimensions investigated, it is hypothesised that trainee teachers recognise greater strengths in their role in activating inclusive practices in everyday teaching, and a stronger criticality linked to a lack of organisational legitimacy. Opportunities also emerge in the construction of local networks and threats arising from the persistent marginalisation of their role within the school. This interpretation can make a significant contribution to rethinking support from a systemic perspective, understood not as an ancillary function but as a structural lever for inclusion (D'Alonzo et al., 2024), while opening up new reflections on the professional profile and identity of teachers (Bruni, 2024) and specialised teachers. In this sense, it is essential to promote in training courses a vision of support as a collegial and cross-cutting action, capable of influencing both the organisational structure and the educational culture of the entire school institution and the wider educational community.

Keywords: Support teachers, roles, systemic vision, training.

The role of the support teacher in inclusive schools: theoretical reflections and research perspectives on co-teaching

Giuseppe Filippo Dettori – Marialuisa Pinna

This paper aims to offer a detailed reflection on the role of the specialised support teacher within the Italian school system, focusing on their function in promoting equitable and inclusive education. The analysis is part of a theoretical framework that values the principles of equity and inclusion, introduced in the Salamanca Declaration (UNESCO, 1994) and subsequently taken up by international educational agendas, particularly in the context of the Sustainable Development Goals of the 2030 Agenda. These principles recognize education as a fundamental human right and a public good, the realization of which requires the active involvement of all actors in the educational context, primarily teachers.

In the Italian context, support teachers play a key role in the implementation of personalized educational pathways, promoting access to quality teaching for all students and acting as mediators between the teaching team, families, and the local community. However, this figure is still often marginalized, subject to stereotypical representations, and perceived as an "appendage" to the student with disabilities, rather than as a resource for the entire class. This hinders the construction of authentic educational alliances and limits the positive impact that could result from the full utilization of their skills.

In this perspective, the bio-psycho-social approach of the ICF (WHO, 2001) provides a useful conceptual model for overcoming a medicalized view of disability, valuing the functioning of the person in relation to environmental factors and promoting self-determination. The school, as a privileged educational context, is called upon to develop environments and practices that make this perspective effective. Teacher training therefore becomes a key issue, so that teachers are able to act in a collaborative and inclusive manner.

Co-teaching is an effective teaching strategy in this regard. Originating in the US (Cook & Friend, 1995), it involves co-planning, co-teaching, and co-assessment by a curriculum teacher and a support teacher. This practice promotes flexible teaching organization, facilitates the management of heterogeneous groups, and contributes to improving the classroom climate and academic performance, while also offering cooperative models to students.

Italian ministerial guidelines (Ministerial Decree of September 30, 2011) and European recommendations (European Agency for Special Needs and Inclusive Education, 2024) insist on the importance of promoting the professionalism of inclusive teachers, based on values such as collaboration, partnership, and teamwork. Despite these guidelines, the reality of Italian schools still appears to be characterized by a lack of a culture of collaboration, hierarchical relationships between teachers, and initial and in-service training that is not always adequate.

Numerous studies in the literature highlight the benefits of collaboration between teachers in terms of professional development, teaching innovation, job satisfaction, and self-efficacy (Hattie, 2015; Meirink et al., 2010). However, several critical issues also emerge: lack of clarity in roles, absence of opportunities for joint training, limited resources, and asymmetry in the perceived skills of curriculum and support teachers. These factors negatively influence the implementation of co-teaching, which is sometimes reduced to a form of assistance or marginal support by the support teacher.

At the international level, recent meta-analyses (Vembye et al., 2023) have highlighted the effectiveness of co-teaching practices in increasing student academic achievement, especially when based on structured collaboration models. In Italy, however, empirical research on the subject is still underdeveloped, within a broader educational research landscape in which the main research trends and models show little inclination to use data collection and tools specific to empirical research (Calvani, De Angelis, Marzano, Vegliante, 2022). Existing studies on co-teaching (Associazione TreeLLLe, Caritas Italiana and Fondazione Agnelli, 2011; Canevaro and De Anna, 2010; Canevaro et al., 2011; Anastasiou, Kauffman and Di Nuovo, 2015; D'Alessio, 2011; Kanter, Damiani, and Ferri, 2014; Norwich, 2015; Ianes and Cramerotti, 2015; Ghedin, Aquario, and Di Masi, 2013; De Anna and Covelli, 2021; Ghedin and Aquario, 2019) highlight critical issues in the implementation of inclusive education and collaboration between teachers.

This paper aims to illustrate the preliminary results of research carried out in a lower secondary school, exploring the potential of co-teaching through a comparison between collaborative practices and traditional teaching approaches. It is also proposed to use INVALSI test data to initiate actions to improve teaching (Castoldi 2025), integrating them with qualitative observations and interviews with teachers, which could offer useful guidance for improving the quality of teaching.

The goal is to contribute to the construction of shared and reliable knowledge on inclusive processes, capable of supporting schools in the transition to more democratic and participatory organizational models. In this perspective, co-teaching is not only a teaching technique but also an educational paradigm that promotes shared responsibility, equality among teachers, the valorization of differences, and student empowerment. Its dissemination requires a cultural change involving the entire school community, initial and in-service teacher training, and a rethinking of educational policies.
In conclusion, inclusion cannot be entrusted to a single professional figure, nor can it be achieved in the absence of authentic collaborative practices. The challenge is to build a school in which the support teacher is recognized as a resource for all, able to contribute to the professional growth of the teaching team, the well-being of students, and the creation of a more equitable society. Only through continuous investment in research, training, and shared reflection will it be possible to transform inclusion from an abstract principle into everyday practice.

Keywords: School inclusion, Support teachers, Co-teaching, Case study

SESSION 8. THE STRATEGIC ROLE OF DATA FOR AN EVIDENCE-INFORMED CULTURE TO SHED LIGHT ON TEACHERS' PROFESSIONALISM IN SOCIAL CONTEXTS

ORGANIZER: UNIVERSITY OF MILAN BICOCCA - INVALSI
COORDINATOR: GIANLUCA ARGENTIN - PAOLO BARABANTI
20TH NOVEMBER: 4.30 P.M. - 6.30 P.M. {AULA MAGNA – RESEARCH 11}

Job Satisfaction and Work Meaning: Teachers in the Italian Labour Market

Ivan Blancato

Introduction

This contribution focuses on a specific segment of the Italian labour market: teachers. This professional group plays a central role due to its size and its crucial function in the education and training of younger generations. Given their importance in society, it is essential to investigate teachers' working conditions and how they perceive them. In particular, job satisfaction offers a valuable lens through which to understand how individuals evaluate their professional condition and, more deeply, the meaning they assign to their role. In short, satisfaction can serve as a synthetic indicator of perceived job quality and thus helps us reflect on this segment of the labour market. In the case of teachers, their subjective perception of work offers important insight into the overall quality of the school system, whose functioning largely depends on those who experience and shape it on a daily basis. Furthermore, teachers' well-being has significant implications for interpersonal relationships and student learning within schools and should therefore be carefully considered.

In this context, the *Meaningful Work* approach provides a relevant framework. As highlighted by Bailey et al. (2019), this literature emphasizes that multiple factors contribute to well-being at work. The approach invites us to look not only at extrinsic aspects of job satisfaction but also at how individuals perceive purpose, coherence, and meaning in their professional lives. The sense each person attributes to their role can strongly influence job satisfaction. This perspective is particularly relevant when studying teachers, as teaching is often associated with a vocational dimension (Cavalli & Argentin, 2010): teachers are often perceived – and describe themselves – as professionals with an educational mission (Colombo, 2017).

Additionally, based on Festinger's (1954) social comparison theory, a large body of research shows that job satisfaction is often evaluated in relative terms: people assess their working conditions by comparing them to those of others. For this reason, it is important to consider that teachers likely compare their position to that of workers in other sectors (Wilkin, 2013). Therefore, when studying teachers' job perception, it is crucial to consider both their working conditions and how these differ from other professions.

Research Objectives and Hypotheses

Following these considerations, our aim is to examine how teachers perceive their working conditions and whether they report being satisfied with their profession. We also seek to explore how job satisfaction varies across different subgroups of teachers. To this end, we compare teachers with other professional groups that are similar in terms of skills and tasks. Since teaching is a profession with a strong vocational component, we believe studying it can help us better understand how the meaning attributed to work can influence subjective well-being and shape labour market experiences.

We hypothesize that teachers report high levels of job satisfaction, both in absolute terms and compared to other workers. In line with previous studies (Argentin, 2021), we expect Italian teachers to report higher average satisfaction than other professionals with similar job characteristics. We also aim to examine the influence of the meaning attributed to one's role on job satisfaction, hypothesizing that this intrinsic component is particularly relevant among teachers and contributes to their high satisfaction levels.

Additionally, we explore job satisfaction across different subgroups of teachers, specifically focusing on how job perception varies by age and contractual status.

Data

To address these research questions, we conducted a secondary analysis of data collected through the Labour Force Survey by the Italian National Institute of Statistics (ISTAT). These data are particularly suitable for the comparative goals of this study: the survey uses a standardized questionnaire that is essentially identical for all respondents, allowing for meaningful comparisons. Moreover, by combining

several survey waves, we were able to build a large dataset of about 250,000 cases, nearly 40% of which are teachers. This large sample allows us to compare specific subgroups and apply multivariate statistical techniques that require substantial case numbers. The dataset includes data from 2014 to 2023.

Methodology

The data analysis was based on the implementation of linear regression models, aimed at estimating the average satisfaction scores among different subgroups of teachers and comparable professions. These models included control variables related to sociodemographic and occupational characteristics, which are potential determinants of job satisfaction. Additionally, we performed robustness checks to identify possible sources of bias in our analyses.

Results

The findings show that, on average, Italian teachers report higher levels of job satisfaction than workers in comparable professions. However, the most prominent result is the strong association between job satisfaction and interest in the work itself.

Among teachers, older age is associated with a decline in job satisfaction—a trend that differs from that observed in other professions. While younger teachers report levels of satisfaction comparable to their peers in other sectors, older teachers show a more marked decline, particularly regarding satisfaction with pay and career prospects.

Finally, employment discontinuity is clearly associated with dissatisfaction about job stability, both among teachers and workers in other fields. A surprising result concerns the relationship between contract type and income satisfaction: among teachers, those with fixed-term contracts report higher satisfaction than those with permanent positions – an inverse relationship compared to what is observed in the other occupations considered.

Keywords: Teachers, Labour Market, Job Satisfaction, Meaningful Work

Peer review as a lever for quality improvement in vocational education and training: roles, competencies, and their assessment

Diego Boerchi – Laura Evangelista – Concetta Fonzo

Introduction

In recent years, the reflection on the quality of vocational education and training (VET) systems has progressively shifted from an accreditation-centered vision—which guarantees compliance with minimum standards, especially organizational and procedural—to a transformative perspective founded on participatory evaluation and oriented toward continuous improvement. In this context, peer review—understood as inter-peer review—has gained a central role and is increasingly recognized at the European level as an evidence-informed tool for promoting quality. The 2020 Council Recommendation of the European Union, which defines the principles of EQAVET (European Quality Assurance Reference Framework for Vocational Education and Training), includes peer review among the preferred methods for implementing evidence-based evaluation, contextualized and grounded in professional comparison.

Unlike other more standardized evaluation mechanisms, peer review is based on interaction among individuals with equal authority and competencies, who are called to observe, discuss, and critically report the strengths and areas for improvement of an organization. It is therefore a process that implies not only procedural knowledge but also significant relational, reflective, and design capacities. It is precisely for this reason that, although formally included in European and national policies, peer review proves difficult to apply without an adequate professional and cultural infrastructure..

Subject, objectives, and research hypothesis

The research presented here, promoted by INAPP, intends to analyze in depth the activities, procedures, and competencies involved in peer review, with reference to quality evaluation processes in regional VET systems. The primary objective is twofold: on one hand, to reconstruct in detail the procedural phases of the process; on the other, to identify the key roles and critical competencies needed so that peer review produces learning effects and organizational improvement.

The underlying hypothesis is that the effectiveness of peer review depends not only on the methodological correctness of the process, but also—and above all—on the ability of the people involved to interpret their

roles in a competent and motivated way. Consequently, it is deemed necessary to move beyond a formalistic view of organizational positions, in favor of an approach that valorizes the actual work performed and the emerging professional behaviors.

Data used

The empirical basis of the research is composed of two main sources: on one hand, a document analysis conducted on national and European references (including CEDEFOP, ARQA-VET, ISFOL, INAPP); on the other, a series of semi-structured interviews with privileged witnesses directly involved in peer review experiences. The interviews, conducted in different Italian regions, have been integrated with the Critical Incidents technique (Flanagan, 1954), in order to identify significant episodes that could highlight the competencies actually mobilized in complex or atypical situations.

Method

The chosen methodological design is based on a combination of task analysis and job analysis, supported by a multi-method approach (Ferrario & Tongiorgi, 1998). Task analysis allowed a systematic description of the activities required in the different phases of the process (planning, self-assessment, peer visit, redesign), while job analysis enabled the articulation of the emerged roles in terms of functions, tasks and competencies, without flattening them onto formal figures.

A distinctive methodological element is represented by behaviour analysis, integrated to capture key behaviors even in the absence of procedural formalization. This made it possible to seize the situated, relational and discretionary dimension of professional action in peer review processes. Interviews were conducted with a semi-structured protocol, adapted to participants' experience level, to guarantee pertinence and reliability of the information collected. The obtained data were then re-elaborated through a thematic analysis process, aimed at building operational models and competency repertoires.

Results

The results of the research are articulated into three main areas.

The first is related to a detailed description of the four phases that constitute the peer review process. For each phase—planning, self-assessment, peer visit, redesign—objectives, expected outputs, operational procedures and main critical issues have been identified. This reconstruction provides the basis for role analysis, making visible the sequence of actions, interactions and decisions characterizing the process.

The second is related to the identification of six recurring key roles in the observed contexts: Director, Quality Coordinator, Facilitator, Research Expert, Improvement Designer, Specific Internal Figures. These roles do not necessarily correspond to formal organizational functions, but represent recognizable sets of operational activities and responsibilities. In contexts with limited resources, multiple roles may be assumed by the same person; however, each role remains critical for the quality of the process.

Mapping of competencies, associated with the different roles, is the third area. For each role a profile was developed, articulated in:

- **Knowledge**: normative, methodological and thematic references (e.g., EQAVET framework, evaluation methodologies, environmental and digital policies);
- **Technical skills and abilities**: data management, logistical organization, meeting facilitation, report writing;
- **Transversal competencies**: communication, collaboration, complexity management, adaptability;
- **Motivational orientations**: openness to dialogue, commitment to improvement, belief in the transformative value of evaluation.

Based on this repertoire, a set of tools was developed for the evaluation and self-evaluation of the competencies required in the different roles involved in peer review processes. The battery includes three distinct but integrated tools: a knowledge test and two self-assessment questionnaires—one focused on technical-specialist competencies (knowledge and skills) and one on transversal competencies.

The knowledge test was designed to measure how deeply a person understands the peer review process, as described in reference documents and experienced in real contexts. The multiple-choice items cover the entire cycle of peer review—from planning to redesign—and address specific aspects such as the purposes of each phase, operational procedures, roles involved, tools to be used and organizational conditions affecting effectiveness. The construction of the test involved expert review to ensure clarity, relevance and balanced coverage of content areas, aiming to provide a tool useful for both evaluative and formative reflection.

The two self-assessment questionnaires, on the other hand, were developed starting from the competency map derived from the job analysis, articulating for each role and each profile dimension (knowledge, ability and transversal competence) a series of descriptive statements referring to observable behaviors.

Respondents, for technical competencies, indicate how capable they are of performing a given task, choosing alternatives from total inability to specialist ability; and for transversal competencies, self-evaluate on a Likert scale how frequently they enact some of the behaviors in their work context that most approximate those they should adopt when participating in a peer review process. The items were formulated to avoid generic or overly abstract statements, instead aiming to anchor evaluations to concrete and recognizable situations.
The tools were subjected to an initial pilot phase, addressed to a sample of teachers, trainers and managers selected both among those engaged in significant peer review roles and those unfamiliar with the process. Preliminary data analysis revealed encouraging results both psychometrically (with good internal consistency of the questionnaires and satisfactory content validity) and in terms of perceived acceptability by participants. The correspondence between the theoretical structure of competencies and the actual responses confirms the solidity of the analysis underpinning tool construction. Looking forward, these tools can be further refined and used for training purposes (e.g., as a basis for designing targeted upskilling paths), for building balanced evaluation teams, and finally as devices for individual and professional reflection oriented toward continuous development.

Keywords: Peer review, competencies, professionalism, evaluation

Focus on Teachers' Well-Being and Motivation in the Light of Evidence-Informed Data

Maria Chirico – Alessandra De Angelis

Introduction

The topic of teacher professionalism is often at the center of national and international debates, especially because students' learning outcomes are considered closely linked to the teaching process.
The teaching-learning process is characterized by an interaction between teacher and student, aimed at personal growth and the development of skills, through the use of multiple teaching methodologies applied within a calm, stimulating, and motivating environment.
The educational phenomenon involves the active co-participation of teacher and student; the teacher plays a supportive and guiding role in the cognitive and metacognitive growth process (Rossini). Therefore, the teacher's role is crucial for achieving effective educational goals, especially considering that many studies highlight the close link between the quality of teaching, teacher training, and students' academic success.
Starting from the principle that the student is at the center of the teaching action, the teacher's role cannot be diminished or underestimated, as it is irreplaceable in the school experience. Teacher professionalism should therefore be encouraged and supported, now more than ever, as it is threatened by widespread social delegitimization, instability and difficulty in accessing the profession, demotivation, challenges in classroom management, complex relational dynamics, lack of family collaboration, overcrowded classrooms, and, in some cases, inadequate technologies and outdated school content.
While the literature agrees on the undeniable mutual influence between teachers and students in the educational relationship (Pontecorvo, 1995; Pianta, 1991), this data appears particularly significant from an emotional-affective standpoint. At this level, the importance of an empathetic and encouraging attitude by the teacher toward students is evident, as is the importance of a positive attitude by students toward their teacher. On the cognitive and metacognitive level, however, the role of this (positive) reciprocity is perhaps less recognized (Rossini).
Today, schools promote various activities aimed at improving the quality and effectiveness of teaching processes. Therefore, teachers are expected to possess increasingly complex skills, such as coordination and management within school organizations, building networks and carrying out innovative experiments involving multiple school stakeholders, initial training and professional development, a holistic approach to the curriculum, and so on. But one may ask how much attention is paid to motivation, self-efficacy, professional well-being, perceived competences, and teaching practices.
The hypothesis of this research stems from the intention to focus on teacher professionalism, since their role and condition have an impact on the teaching and learning they promote.

Thus, the research aims to move beyond an approach based on stereotypical models and idealized conceptions of teaching and teachers, trying instead to explore the elements that characterize teachers' professional identity, self-perception, motivation, psycho-physical well-being, and the level of perceived competence.

Research Subject, Objectives, And Hypothesis

The research hypothesis targets a sample of teachers from all school levels (pre-school, primary, lower secondary, and upper secondary) to whom a questionnaire was administered during the 2024/2025 school year.

The survey includes four areas:

1. Analysis of respondents' data, such as age, gender, academic qualifications, school level, region, years of teaching experience, subject taught, and institutional roles held;
2. Passion/motivation/interest for the teaching profession;
3. Self-perception of their work and perceived skills;
4. Teachers' psycho-physical well-being and environmental conditions.

Respondents were asked to answer over 40 items anonymously.

The objective is to highlight the importance of "other factors" that affect the quality of teachers' work — namely, understanding their lived experience, needs and expectations, relationships within the school community, and their motivation and passion for teaching.

Data Used

300 questionnaires were administered to teachers from various school levels across different regions of Italy, from North to South.

Method Or Approach

The method used for this quantitative survey was interviews conducted through Google Forms.

The direct administration of the questionnaire to teachers allowed an evidence-based approach, focused on data relating to teachers' motivation and perceived efficacy, the competences currently required for teacher professionalism, and the impact of more equitable school environments.

Results Or Discussion

The results of the questionnaire analysis were noteworthy: over 80% of respondents were women and held university degrees; more than 40% were aged between 50 and 60 years; over 70% taught in upper secondary schools; around 60% worked in Southern Italy. The distribution by years of teaching experience, from 5 to 30 years, was fairly even, with about 15% in each experience range considered.

Regarding the subject taught, around 80% of teachers taught their own subject, while the remaining were involved in support teaching.

The sample examined revealed that 80% of subject teachers teach, for approximately 50% of their time, the subjects assessed by the Invalsi tests, namely Italian, Mathematics, and English.

Within the school, 50% of these teachers serve as Class Council Coordinators, nearly 15% act as department coordinators, subject coordinators, or PCTO (school-to-work programme) tutors. The remaining percentages concern orientation tutors and advisors, digital facilitators, civic education coordinators, site managers, and school well-being and health coordinators, while 20% hold no additional roles.

90% of respondents reported a strong passion for their profession, while only 10% do not feel the same commitment. The motivation to teach one's subject remains high, exceeding 87%. Around 70% of teachers report being more than satisfied with the feedback they receive from their students regarding their work; 60% are very satisfied with the level of recognition from students' families; 50% receive highly positive recognition from colleagues on the Class Council, whereas over 65% report being almost completely dissatisfied with the level of recognition from the Head Teacher and their team.

60% of teachers consider themselves highly satisfied with the effectiveness of their work with students.

Over 80% report having experimented with good teaching practices in their classrooms, and over 43% are very satisfied with the results achieved.

Only about 36% perceive a positive level of well-being within their school, and 40% report a positive atmosphere among their school colleagues, whereas 35% state that the Head Teacher and their team establish functional relationships with the teaching staff and believe that strong action is needed to improve communication and collaboration among the various stakeholders.

Over 75% of teachers consider the presence of an internal mental coach essential to support the teaching profession, also due to the fact that more than half of the sample have experienced conflictual situations with colleagues, and about 25% have had at least one episode of burnout during their career.

Additionally, more than 80% believe that the physical environment in which they work has a significant impact on teacher well-being, yet only 25% consider the internal and external school spaces adequate.
Over 80% believe that the "teacher status" is poorly recognized by society.
Only 10% report working 18 hours per week, while 35% work over 30 hours weekly.
More than 70% believe they have the necessary skills to perform their job, and over 79% consider investment in teacher training to be very important. More than 90% expect the Ministry of Education (MIM) to invest more in teacher training, particularly in psychological-pedagogical areas, subject-specific training, and artificial intelligence.
Over 95% of the sample believe their economic treatment is absolutely inadequate, yet around 80% confirm their choice to teach and believe that benefits/incentives (such as bonuses, expense reimbursements, meal vouchers, financial and cultural incentives, and psychological support) are necessary to stimulate teacher motivation.

Keywords: Teacher professionalism, motivation, well-being, evidence-informed

Assessment Practices and Opinions of English as a Foreign Language Teachers: An Exploratory Study Based on Data from the INVALSI Teacher Questionnaire

Francesca La Russa

Introduction

Due to their role as a co-protagonists in the educational process and their strategic function in facilitating and promoting learning, teachers hold significant educational and political importance. For this reason, beside the national tests administered to students, INVALSI also asks teachers of the sample classes to complete the Teacher Questionnaire.
The Teacher Questionnaire is a research tool designed to investigate teachers' opinions on the INVALSI tests and to identify school-context and teaching-related factors that may influence student performance (INVALSI, 2014a).
Structured on three levels, the Questionnaire examines:

- at the individual level, teachers' attitudes toward external assessment, their professional experience, and demographic characteristics;
- at the classroom level, the methodologies and teaching practices adopted;
- at the school level, perceptions of organizational well-being.

Several studies (Pastore & Freddano, 2017; Leggi et al., 2020; Barabanti, 2018; Le Rose & Falzetti, 2021) have analyzed responses to the Questionnaire provided by Italian and Mathematics teachers. With the introduction of the national English test in the 2017-2018 school year, the Questionnaire has also been administered to English language teachers. However, to our knowledge, research on responses provided by English teachers remains limited.

Aims

This study aims to provide an overview of the assessment-related opinions and practices of English language teachers in the fifth year of primary school (Grade 5), the third year of lower secondary school (Grade 8), and the final year of upper secondary school (Grade 13), examining potential similarities and differences across the three educational levels.
The analysis focuses on two main aspects:
a) teachers' opinions on the external assessment proposed by INVALSI;
b) classroom assessment practices.

Data and Methods

The data come from the INVALSI Teacher Questionnaire administered in the 2022–23 school year to 761 English language teachers in Grade 5, 160 teachers in Grade 8, and 788 teachers in Grade 13, coming from various regions across Italy.
As this is an initial exploratory analysis, the focus will be on descriptive statistics related to the following items:

- questions 1 to 4 from the Grade 5 teacher questionnaire and questions 1, 2, and 4 from the Grade 8 and Grade 13 questionnaires, which concern teachers' opinions on the INVALSI tests, how teachers

prepare students for the test, and their familiarity with the assessment framework—namely, the Common European Framework of Reference for Languages (CEFR, Council of Europe, 2001; 2020);
- questions 7 and 8, which focus on classroom assessment practices and how student errors are used pedagogically.

Results

This exploratory analysis offers a general overview of the assessment-related opinions and practices of English language teachers across different school levels. Overall, the external assessment proposed by INVALSI is perceived positively by teachers at all three educational stages. This finding is particularly significant if assessment is viewed as a system: a mosaic of diverse yet complementary components in which external assessment is not in opposition to classroom assessment, but rather serves as a resource that can enrich and complement existing practices. In this perspective, teacher engagement is essential to ensure that data from external assessments are used effectively.

The results show that teachers do prepare their students for the INVALSI tests, mainly by using the resources provided by the Institute. This allows students to become familiar with the structure and administration of the test, which can help reduce test-related anxiety. The choice to use official resources appears justified, given the extensive validation and calibration work conducted by subject matter experts and psychometricians during test development. This makes the INVALSI materials more reliable than unofficial simulations. However, it is important to emphasize that "teaching to the test" can be effective only when it aims to develop transferable skills and competences that go beyond merely passing the test. In contrast, an approach that prioritizes repetitive and mechanical exercises at the expense of curricular content may prove counterproductive (Pozio, 2014).

Regarding the CEFR, while most teachers report having at least skimmed the document, only a minority state that they have read it thoroughly. Considering that the CEFR serves not only as the framework for the INVALSI tests but also as the foundation for the learning objectives in the national curriculum, there is a clear need to strengthen teachers' familiarity with and informed use of this key reference tool in their daily practice.

As for classroom assessment practices, traditional methods remain predominant across all school levels, whereas more innovative approaches, such as the use of digital technologies or real-world tasks are still not enough widespread, although there is some indication of growing interest at the higher educational levels.

Error is widely recognized by teachers as a meaningful component of the teaching and learning process. A shared view emerges of its formative value, as it is frequently used to encourage reflection and collaborative learning in the classroom. While student errors could also be leveraged by teachers to reconsider their own methodological choices, analyzing student errors for the purpose of self-evaluating one's teaching approach is not yet a common practice. Finally, although the New National Guidelines for the curriculum promote an understanding of assessment as a collegial process grounded in professional dialogue, the practice of reflecting on student error through peer discussion remains relatively uncommon.

Keywords: INVALSI Teacher Questionnaire, English as a Foreign Language, teaching practices, assessment

Student Participation, School Climate and Educational Governance: A National Multi-Stakeholder Survey in Upper Secondary Schools

Giulia Gabriella Pastori – Valentina Pagani – Mara Soncin – Alessandra Maiorano

Introduction

International research on school improvement acknowledges the strategic relevance of student participation in promoting well-being, inclusion, and learning (Mitra, 2018; OECD, 2024). However, in the Italian context, systematic investigations capturing the perspectives of students, teachers, and school leaders on engagement, active citizenship, and school climate remain scarce. The project 'Costruire Futuro Insieme 2' (CFI2), promoted by ActionAid Italy and the School and Democratic Citizenship Laboratory at the University of Milano-Bicocca, addresses this gap by conducting a national multi-stakeholder survey. The study's originality lies in the integrated and coherent construction of three parallel questionnaires that share key analytical dimensions—such as perceptions of school climate, representations of participation and civic education, and forms of engagement—enabling comparative analyses across roles.

Research focus, objectives and hypotheses

The research aims to investigate, from the perspectives of students, teachers, and school leaders, the extent and quality of participatory practices; to explore how participation intersects with individual (motivation, agency), professional (teacher identity, leadership), and institutional (relationships, governance) factors; and to provide empirical evidence supporting a democratic rethinking of school organization. The general hypothesis posits that the limited student participation in teaching, school life, and territorial contexts is rooted in systemic and cultural conditions, and that its reinforcement could enhance equity and cohesion within the education system.

Data used

The sample includes 2,023 students from 17 Italian regions, representing all types of upper secondary schools (lyceums, technical, vocational); 543 teachers from 13 regions, with an average age of 48.6 and a mean of 17 years of teaching experience; and 16 school leaders, the majority of whom have fewer than ten years of experience. The sample was stratified by macro-region and school track, ensuring balanced contextual variables. Data collection took place between November 2024 and March 2025.

Method

The questionnaires were developed using nationally and internationally validated tools, supplemented with original items and experimental techniques, such as factorial vignettes, to elicit perceptions and judgments in realistic scenarios. The shared dimensions across the three tools allow for a triangulated analysis of participatory dynamics. The student questionnaire included dimensions such as school climate and relationships (Multidimensional School Climate Questionnaire – Grazia & Molinari, 2020); perceived teaching style (Ferguson et al., 2015; Nanwani, 2019); school engagement in its emotional, behavioral, and cognitive forms, and agentic engagement (Mameli & Passini, 2018); psychological capital (PsyCap, Mazzetti et al., 2018); and civic participation. The teacher questionnaire examined professional identity (Velasco et al., in press), organizational climate (MSCQ), instructional practices promoting student agency (Ferguson et al., 2015; Nanwani, 2019; OECD, 2024), work engagement and relational dimensions (Avanzi et al., 2013), and self-efficacy (NTSES). Civic education was also addressed through items on goals, content, and practices. The school leader questionnaire was based on a robust theoretical framework on distributed leadership (Spillane, 2005) and validated scales by Hulpia, Devos, and Rosseel (2009) to analyze team management, staff involvement in decision-making, and leadership coherence. It also incorporated items from literature on school governance (Maslowski et al., 2016; Agasisti, Rossi & Soncin, 2021) and the World Management Survey (Di Liberto, Schivardi & Sulis, 2015).

Limitations

Despite the methodological soundness and the diversity of the sample, the study has some limitations. School participation was voluntary, which may have introduced self-selection bias favoring schools already attentive to participation issues. Although the territorial distribution was stratified, responses were more concentrated in Northern and Central regions, with lower representation from Southern Italy and the Islands. The limited number of participating school leaders constrains the possibility of broader comparative analyses, though it offers relevant qualitative insights.

Findings and Discussion

The data analysis reveals a complex and nuanced landscape, marked by partial convergences and significant misalignments among the three perspectives explored. Student participation, though widely acknowledged as important by all stakeholders, appears to be characterized by a persistent gap between principles and actual practices. A particularly salient finding is the divergence between teachers' and principals' perceptions of how student participation and agency are promoted. Teachers generally recognize the importance of inclusive and participatory learning environments and report using methods that value student voice. However, their practices tend to remain within the didactic sphere and are often limited by structural and cultural constraints. On the other hand, school leaders, while rhetorically embracing distributed leadership and student protagonism, seem to implement these principles only weakly and inconsistently. Students are often perceived as marginal actors within school governance, with involvement confined to symbolic or consultative roles. From the students' point of view, participation opportunities exist but are frequently perceived as formal, episodic, or ineffective. While student councils are widespread, many students report not feeling heard or not knowing their representatives. The prevailing school experience remains largely transmissive and hierarchical, with limited space for initiative, responsibility, and co-design. Another critical issue concerns school–community relations. Students report greater exposure to initiatives related to career guidance and higher education than to civic engagement or collective action. Principals often value partnerships primarily for organizational or project-based purposes, rather than as opportunities for democratic school development.

Conclusions
Student participation cannot be reduced to a formal or bureaucratic issue but must be recognized as a key indicator of the democratic quality of schools. The data call for restoring coherence between value declarations and everyday practices, recognizing participation as a strategic lever for building more equitable, sustainable, and future-oriented schools. The survey reveals a scenario in which students, despite being included in multiple participatory mechanisms, still play a predominantly passive role in school life. Their voice often struggles to influence decision-making processes, and opportunities for genuine agency remain limited or symbolic. In this context, engagement, agency, and school participation emerge as essential dimensions not only for recognizing students' subjectivity but also for constructing meaningful learning environments, relational well-being, and motivation to study. The CFI2 project can also be considered a pilot study, useful for developing integrated indicators and tools for systematic monitoring. The findings offer a foundational knowledge base for informing school policies and training pathways. Specifically, the data can support the redesign of initial and in-service training for teachers and school leaders, so that they are equipped to recognize and promote participation as a core pedagogical and organizational dimension. Using data in a formative way may help bridge the gap between ideals and practice, strengthening awareness of each actor's role in shaping a democratic, inclusive, and high-quality educational system.

Keywords: Student participation, School climate, Agency, Educational governance

SESSION 6. EVALUATING THE IMPACT OF EDUCATION POLICIES: COUNTERFACTUAL APPROACHES AND EVIDENCE FOR BASIC SKILLS IMPROVEMENT

ORGANIZER: EUROPEAN COMMISSION, DIRECTORATE GENERAL FOR EDUCATION, YOUTH, CULTURE AND SPORTS

COORDINATOR: MARCO MONTANARI - DANIELE VIDONI

21ST NOVEMBER: 8.30 A.M. - 10.45 A.M. {AULA 5 – RESEARCH 12}

The Impact of School Canteens on Academic Performance: Evidence from Primary Schools

Federica Dalponte

The international community is increasingly recognizing school meal programs as a transformative strategy to simultaneously address child nutrition, education, health, and social development, with initiatives like the School Meals Coalition, aiming to ensure that every child receives a school meal by 2030. There is a relevant strand of literature examining the impact of school meals on academic performance, which generally finds a positive effect (e.g., Cohen et al., 2021; Gordanier et al., 2020; Leos-Urbel et al., 2013; Ruffini, 2022; Schwartz & Rothbart, 2020). However, these studies are mainly US-based and predominantly investigate the impact of universal free school meals versus income-based meal pricing strategies. This paper contributes to this literature by examining the impact of the presence (versus absence) of school canteens in Italian primary schools on academic performance, which might entail potentially broader mechanisms beyond purely economic ones. Italy represents an interesting setting for this study, as it has recently given increased attention to the topic, with an investment of €1.07 billion from the National Recovery and Resilience Plan to construct or renovate canteens in at least 1,000 school buildings by 2026. One of the channels through which the presence of school canteens might positively influence academic performance is by being an important condition for the extension of school day in the afternoon, which in turns provides additional education activities, more opportunities for peer interaction and structured time for homework. Therefore, this paper also contributes to the literature on the impact of extended school time on academic performance, which generally finds a positive effect (e.g., Battistin & Meroni, 2016; Bellei, 2009; Cattaneo et al., 2017; Huebener et al., 2017; Lavy, 2015). The presence of school canteens might also impact positively on students' academic performance through the provision of at least one proper meal per day. Therefore, this paper is also related to the literature examining the impact of nutritional intake quality on academic performance, which generally finds a positive effect of better dietary quality (e.g., Anderson et al., 2018; Belot & James, 2011; Burrows et al., 2017; Florence et al., 2008; Jyoti et al., 2005). Indeed, proper nutrition plays a crucial role in supporting academic performance by influencing physical development, cognitive functions, and behavioral aspects.

Methodologically, the paper makes use of data on the presence/absence of canteens in primary schools from the Italian Ministry of Education, for academic years from 2017-18 to 2022-23, and individual level INVALSI data on standardized test scores for Italian, Math, and English in second and fifth grade, for the same academic years. Exploiting the temporal variation in school canteen construction across different schools and periods, a staggered difference-in-differences methodology with repeated-cross sections is employed, both with linear regressions with two-way fixed effects and Callaway and Sant'Anna estimator, with school and year fixed effects and individual level covariates to control for potential compositional changes. Relying on a sample of more than 800 thousand observations, the study finds that the presence of a school canteen has a positive and statistically significant impact on standardized test scores for Italian (0.02 standard deviations from the average performance in the overall INVALSI tests population) and English listening (0.05 standard deviations) in fifth grade, while no effect is observed in second grade. These results are robust to parallel trend tests, both with event study design and placebo test artificially assigning treatment one year earlier than it actually occurred. Robustness checks testing whether other concurrent events or trends might have influenced the observed academic performance also ensure that the estimated effects are specifically attributable to canteen construction and not to alternative simultaneous factors. The study uncovers heterogeneous treatment effects based on student characteristics. Students speaking a language

other than Italian at home have a more positive effect on English listening score, while their effect on Italian is as positive as that of students speaking Italian at home, with Italian being a subject in which students speaking a different language at home usually score worse. Students with delayed academic progression, who typically underperform compared to their on-time peers, show even larger improvements in English listening score. Gender dynamics emerge as another interesting dimension, with the presence of a school canteen favoring male students in Italian, whose score in this subject is usually worse than female students. Importantly, no heterogenous treatment effects are found depending on students' socio-economic and cultural status, which typically represents a relevant source of disparity in academic performance, favoring more advantaged students. For what concerns the underlying mechanisms, the paper tests the mechanism of increased school time through a mechanism blocking design. The findings suggest that the effect on Italian performance might be entirely explained by increased school time in the afternoon due to the possibility of having lunch at school thanks to the presence of the canteen. Conversely, the positive effect on English listening scores persists even when the increased school time mechanism is not at play, meaning that other mechanisms also explain this positive impact, which however need further investigation.

Keywords: School canteens, primary school, impact, academic performance

Computing Education and Computational Thinking: A Counterfactual Analysis Using ICILS 2023 Data

Riccardo De Alessandri – Giovanni Lari

Introduction

In recent years, the promotion of computational thinking has become an explicit priority of European education policies. The Council of the European Union Resolution on a European Education Area (2021/C 66/01) emphasizes the importance of strengthening basic and advanced digital competences, integrated into school curricula, to address the challenges of the digital transition. In particular, it is recognized that educational success and inclusion also depend on future-oriented training that enhances key competences such as computational thinking, considered essential for active citizenship and preparation for the world of work (Council of the European Union, 2021).

Computational thinking is defined as "the ability to recognize aspects of real-world problems that are appropriate for computational formulation, and to evaluate and develop algorithmic solutions to those problems so that the solutions could be operationalized with a computer" (Duckworth & Fraillon, 2025, p. 38). It is a set of cognitive skills that enable individuals to understand, analyze, and solve problems systematically and algorithmically. Within ICILS 2023, the CT construct is structured into two main dimensions: the first concerns the understanding and representation of problems (including digital systems and data), while the second relates to the design and implementation of solutions (through algorithms, programs, and interfaces). This definition reflects a broad view of CT as a key competence for addressing complex challenges in digital and non-digital environments (Duckworth & Fraillon, 2025).

Schools can promote these competences through the teaching of computing, either as a standalone subject or integrated transversally across other disciplines. However, the concrete implementation of computing education varies widely across educational contexts: not all schools offer computing as a separate subject, and even when they do, the hours allocated, pedagogical approaches, and teacher training may differ significantly. In this context, it becomes relevant to understand to what extent, controlling for individual student characteristics, the mere presence of computing as a standalone subject is associated with higher levels of computational thinking.

The ICILS 2023 (International Computer and Information Literacy Study) provides a particularly valuable data source for analyzing this relationship, as it not only measures standardized student competences in CT but also collects information from school principals and ICT coordinators about school characteristics, policies adopted, and the availability of educational and digital resources. These data contribute to the construction of detailed national profiles that help to understand how computational thinking is integrated into different education systems, both at the curricular and infrastructural levels. This enriched information enables more robust and contextualized analyses of student outcomes and facilitates the evaluation of the

impact of school policies and the conditions of educational provision on the development of digital competences.

Research object and hypothesis

This study aims to estimate, through a counterfactual approach, the average impact of having computing as a distinct subject in lower secondary schools (grade 8) on students' scores in the computational thinking (CT) scale. The hypothesis is that students attending schools offering computing as a standalone course achieve, on average, higher results compared to students with similar observable characteristics who do not have access to the same curricular provision.

Data

This analysis uses ICILS 2023 microdata released by the IEA (International Association for the Evaluation of Educational Achievement), restricted to students from European participating countries. The outcome variable consists of scores on the computational thinking scale, structured into four proficiency levels defined by predetermined thresholds, allowing classification of students according to their degree of mastery of knowledge and skills assessed by the test.

The treatment variable is derived from students' responses to a specific questionnaire item: "Do you study [computing, computer science, information technology, informatics, or similar] in the current school year?". Students could indicate whether they studied these subjects as a separate discipline, within other subjects, or not at all. For this study, students who reported studying computing as a standalone subject are considered "treated," while the other two categories together form the comparison group.

The impact estimation model includes several individual control variables. Age is included to account for possible marginal differences in cognitive development and schooling, despite being relatively homogeneous among students enrolled in the same grade. Gender is controlled for because, as reported in the ICILS International Report, there are average score differences between male and female students: differences that are generally small and not always statistically significant at the national level but potentially relevant as confounding factors.

The choice to control for migration background and language spoken at home is based on ICILS 2018 and 2023 evidence, which shows systematic differences in CT scores associated with these factors. On average, students without a migration background achieve higher scores (approximately 34 points difference), as do those who most frequently speak the test language at home (about 52 points difference), with statistically significant results in most countries. Similarly, socioeconomic status is considered a crucial dimension for control. ICILS data document significant differences in CT scores related to parental education and occupation, as well as the availability of cultural resources at home. In particular, mean score differences of 37 points were observed for students with at least one parent holding a university degree, 49 points for those with parents in highly qualified professions, and 61 points for students reporting more than 26 books at home.

Finally, the variable measuring expected educational attainment is included as an indicator of school motivation, as previous studies (ICILS International Report) have shown that higher confidence and educational aspirations are associated with higher performance in digital competences and, plausibly, also in computational thinking. Including these variables helps to reduce selection bias and to more accurately estimate the treatment effect (Fraillon et al., 2023).

Method

The average treatment effect on the treated (ATT) is estimated through a quasi-experimental design based on propensity score matching. This technique allows comparing students who attend a standalone computing course with similar students without access to the same curricular provision, thereby reducing selection bias due to non-random assignment. Specifically, a radius matching approach is adopted, defining a maximum acceptable distance between the propensity scores of treated and control students. To test the robustness of estimates, the analysis will be replicated using three different caliper values (0.01, 0.05, and 0.1). Using a small caliper improves covariate balance, while larger thresholds help retain an adequate number of observations and increase statistical power. Matching quality is assessed through covariate balance checks, analysis of the distribution of propensity scores, and calculation of matching rates.

Expected results

For exploratory purposes, a preliminary analysis was conducted using propensity score matching with a 0.01 caliper to verify the consistency of the expected association between the presence of computing as a standalone subject and computational thinking scores. The propensity score was estimated through logistic regression on the main individual control variables. Results showed that before matching, the treated group had a mean score about 29 points higher. After matching, the difference decreased to about 24 points, while

remaining statistically significant. These values are purely exploratory and serve to guide the design of the study, without being interpreted as conclusive effect estimates

Keywords: Computational thinking, computing education, Counterfactual evaluation, ICILS

The effect of school district consolidation on student achievement and staff absenteeism

Letizia Gambi – Kristof De Witte

This paper investigates the impact of school district consolidation on two key school outcomes in Flemish primary schools: student achievement (as measured by standardised test scores in Dutch language and mathematics) and staff absenteeism (as measured by the total number of staff absence days per school year). Leveraging the staggered rollout of a large-scale consolidation policy[1] initiated in 2015 by the network of Flemish Catholic schools, the analysis provides quasi-experimental evidence using an event-study difference-in-differences (DiD) framework.
To identify causal effects, this paper applies a recent DiD estimator developed by (de Chaisemartin & D'Haultfœuille, 2024), which accommodates staggered, non-binary, and non-absorbing treatments. The dataset comprises thirteen years (2012–2024) of panel data covering standardised test scores and administrative records from over 1,000 primary schools. The policy intervention is defined by administrative consolidation events that is, increases in the number of schools governed by a single school board, resulting from the merger of two or more districts.

Specifically, the paper addresses the following research question:

RQ4: *"What is the effect of school district consolidation on school outcomes, specifically student achievement and staff absenteeism? Do these effects vary by treatment and district characteristics?"*

Several theoretical mechanisms underpin the expected effects of consolidation. On one hand, insights from the school consolidation literature (Beuchert et al., 2018; Hanushek et al., 2004) and the corporate mergers and acquisitions literature (Steigenberger, 2017) suggest that disruption effects — such as increased uncertainty, greater organizational complexity, and higher coordination burdens — may negatively affect staff well-being and, eventually, interfere with the student learning environment. On the other hand, greater organizational effectiveness may arise from enhanced staff professionalisation, and increased policy-making capacity through synergies created in managerial and instructional practices (Andrews et al., 2002; Dodson & Garrett, 2004). Specifically, organizational effectiveness may improve through enhanced staff professionalisation and increased policy-making capacity, driven by synergies in managerial and instructional practices (Andrews et al., 2002; Dodson & Garrett, 2004). Additionally, economies of scale may promote greater cost-efficiency, while shared professional expertise and pooled resources can enhance educational quality (Andrews et al., 2002; Howley et al., 2011; Leach et al., 2010). Based on the aforementioned mechanisms, we formulate the following hypotheses, one for each outcome of interest:

"It is a priori unknown whether consolidation has an effect on student achievement."

"Staff absenteeism is expected to rise in the short term due to transition-related stress but decline in the long term as organizational efficiency improves."

[1] Over the past three decades, education systems in most OECD countries have experienced a significant decentralization of (education) governance, granting increased autonomy to local authorities, school districts, and individual schools to better meet heterogeneous and local needs (Burns & Koester, 2016). As a result, the school district level has taken on an increasingly prominent role in many education systems – not only in traditionally decentralized contexts such as the United States, but also in countries like the Netherlands, England, and Belgium (Honingh et al., 2020). In these systems, decision-making authority, policy development, and responsibility for educational quality have progressively shifted to school boards (Hooge et al., 2019; Hooge & Honingh, 2014).[1] At the same time, public sector reforms inspired by New Public Management have driven a parallel trend toward consolidation, aimed at achieving economies of scale and enhancing managerial professionalism (Alonso et al., 2015), including at the school and school district levels (Jarl et al., 2012).

Our results show that consolidation had null effects on the achievement of students whose districts consolidated. At the same time, we observe a substantial and statistically significant increase in staff absenteeism following district consolidation – with each unit increase in consolidation leading to a statistically significant rise of 9.5 absence days per year, roughly a 4% increase relative to baseline. These effects intensify over time, peaking at approximately +70 days by year four (a 27% increase), particularly in districts experiencing repeated consolidations (+10.19 days, or a 4% increase), whereas districts consolidating only once show no statistically significant increase in staff absences. Effects are also stronger in larger consolidations (i.e., involving more schools), with an average increase of +45.85 days per year (an 18% rise), and in districts with more complex structures prior to treatment (+29.10 days, or a 12% increase). Furthermore, staff absences increase significantly over time in larger districts (whether defined before or after consolidation), but not in smaller ones. These findings are robust to a range of empirical checks, including staggered treatment designs, placebo tests, and sensitivity analyses with alternative specifications.

On the basis of our results, we concluded that school district consolidation had no statistically significant effect on student achievement in either Dutch language or mathematics, even up to four years after the consolidation event. These null effects are consistent across different specifications and levels of treatment intensity, suggesting that consolidation does not meaningfully alter student learning outcomes in the medium run. By contrast, consolidation led to a statistically significant and sustained increase in staff absenteeism. These effects did not decline over time, contradicting the expectation of long-term improvement through greater organisational efficiency. Instead, absenteeism effects intensified in schools located in districts that experienced larger or repeated consolidations, had more complex structures, or were already large before treatment.

Further heterogeneity analyses revealed that the negative effects on staff attendance were concentrated in districts offering all levels of education (from pre-primary to secondary), while smaller and simpler districts experienced minimal or no effects. These patterns suggest that organisational strain, rather than efficiency gains, dominated the early to medium-term effects of consolidation.

In summary, this paper concludes that school district consolidation does not affect student achievement in language and mathematics at the end of primary education, but significantly increases staff absenteeism – particularly in large, complex, or repeatedly consolidated districts – indicating that the organisational disruptions introduced by structural reforms can have important implications for staff well-being and, potentially, for long-term educational quality. Even in the absence of medium-term effects on student achievement, the negative impact on staff attendance may, in the long run, in turn negatively affect test scores through multiple avenues. Specifically, the importance of teachers for student achievement (e.g., Chetty et al., 2014; Staiger & Rockoff, 2010) and the detrimental effects of teacher absenteeism on student outcomes (Carrizosa & De Witte, 2024) have been extensively documented in the literature. It is thus important to monitor absenteeism, especially among teachers, and future research should explore whether the observed rise in staff absenteeism translates into negative impacts on student outcomes in the longer term.

Keywords: School district size, Consolidation, Standardised tests, Staff absenteeism

Impact of Social Media on Students' Academic Performance

Giovanni Lari – Riccardo De Alessandri

Introduction

This study aims to assess the effect of social media use on students' performance, one of the groups that makes the greatest use of these platforms (Aitken et al., 2024). Performance is measured through the score obtained in the Computational Thinking (CT) test administered as part of the international ICILS study, promoted by the IEA (International Association for the Evaluation of Educational Achievement).

The analysis is based on data from students in European Union countries that participated in ICILS, for a total of approximately 90,000 observations. The effect of social media use is estimated using a counterfactual approach, applying a Propensity Score Matching method with caliper.

Research Scope and Objectives

Students' performance has been deteriorating in recent years (OECD, 2023). In light of the most recent research in the field of education, there is an ongoing debate about which factors are relevant in driving this phenomenon. Since this trend affects students worldwide, in a nearly uniform way across different countries, it is necessary to address the issue from an aggregated perspective, seeking common answers that concern the student population as a whole.

The COVID-19 pandemic is certainly one of the main factors to consider; research is still investigating what its impact on students has been and will be (Betthäuser et al., 2023). An event of such magnitude has drastically changed students' habits and school life. For a significant period, many of them were forced to live in an asocial context, experiencing isolation at an age when adaptation and transformation are entirely natural processes.

Social media represents another global phenomenon that has significantly transformed human habits over the past decade. Beyond the time spent in front of screens, the advent of social media has also changed the way people access information and communicate. The potential implications of social media use are numerous, and probably an entire volume would not suffice to list them all. Social media use today encompasses a wide range of digital behaviors, from instant messaging to passive content consumption, and increasingly to the activity of scrolling. In particular, "infinite scrolling," meaning the ability to consume content without limit or interruption, has been associated with a decline in prospective memory—that is, the capacity to remember to perform an action at a future time (Meinhardt et al., 2025).

In recent years, research has increasingly focused on understanding the effects of intensive social media use, highlighting numerous negative aspects, such as the decline in memory and concentration skills. Therefore, it is expected that these issues may influence students' CT test scores, which rely on complex cognitive processes such as logical organization, abstraction, problem solving, and the manipulation of algorithmic structures (ICILS, 2018).

The objective of this research is to empirically evaluate the impact of social media use on students' academic achievement, contributing to the body of studies investigating the progressive decline in their school performance.

Data

The analysis relies on data collected in the context of the international ICILS 2023 study (International Computer and Information Literacy Study), promoted by the IEA (International Association for the Evaluation of Educational Achievement). The study aims to measure digital competencies and computational thinking among lower secondary school students (grade 8) across a broad range of countries. In particular, ICILS provides detailed data on student performance in standardized tests, as well as contextual information gathered through questionnaires administered to students, teachers, school leaders, and families.

For the purposes of this research, the ICILS 2023 database has been restricted to EU member countries only, in order to ensure greater institutional, cultural, and educational homogeneity in the analyzed sample. The resulting subsample includes approximately 90,000 students, for whom both the CT test scores and relevant sociodemographic and behavioral information are available to estimate the impact of social media use.

This dataset therefore offers a robust and comparable context to empirically analyze the relationship between intensive social media use and computational skills in the school environment.

Method or Approach

The analysis aims to estimate the impact of intensive social media use on students' performance using a counterfactual approach. The principle behind the counterfactual approach is to compare the performance

of students who make intensive use of social media with that of students who are similar in all other observable characteristics but do not use social media intensively. In particular, the Propensity Score Matching (PSM) method with caliper is applied.
Treated students are matched to control students with similar propensity scores, within a predefined radius (caliper), in order to minimize selection bias and ensure a valid comparison between groups.
The control variables used in estimating the propensity score include:

- Age
- Gender
- Motivation
- Migration Background
- Language spoken at home
- Socioeconomic status index (constructed based on the number of books at home, parents' education level, and parents' occupation)

These variables were selected as relevant for performance on the ICILS 2023 Computational Thinking test. The goal is to estimate the net effect of intensive social media use while controlling for potential confounding factors that could compromise the causal interpretation of the results.

Results

The analysis aims to estimate the impact of intensive social media use on students' performance in the domain of Computational Thinking (CT), using ICILS 2023 data and a counterfactual approach based on Propensity Score Matching (PSM) with a caliper of 0.05. At present, the results should be considered preliminary, as they are based on the use of only one plausible value from the CT test, without yet integrating the five plausible values provided by the standard methodology.
Three separate models have been estimated, each defining the treatment group differently based on the frequency of social media use. The relevant survey question measures how often students use social media while doing schoolwork, with five response categories: never, almost never, sometimes, often, very often.
In the first model, treatment is defined as frequent or very frequent use (often and very often), compared to all other categories. This approach has the advantage of preserving the entire available sample. The estimated ATT is −13.42 points on the CT score, with a t-statistic of −12.03, suggesting a statistically significant relationship between frequent social media use and poorer CT performance.
The second model adopts a more selective definition: it compares students who use social media very often with those who report never using it. This choice reduces the sample size (29,306 observations) but allows observation of the potential effect of extreme digital exposure. In this case, the ATT is −29.30 points, with a t-statistic of −13.92, indicating a stronger association between very frequent use and lower performance.
The third model slightly broadens the control group, including students who report never or almost never using social media, maintaining a good balance between conceptual contrast and sample size (34,406 observations). The estimated average effect is −28.60 points, with a t-statistic of −18.66, confirming the trend observed in the other models.
In summary, the results suggest a negative relationship between intensive social media use and computational thinking skills, consistent with the hypothesis that frequent digital behaviors may interfere with the cognitive functions required to solve complex problems. However, it is emphasized that the analysis so far has used only one plausible value, and it will be necessary to replicate and refine the estimates by integrating all five values to ensure further statistical reliability of the results.

Keywords: Social Media Use, Computational Thinking, Academic Performance, Propensity Score Matching

Neoliberal and Democratic school-level policies: Global distribution and consequences for achievement inequalities

Moris Triventi – Andrea Pietrolucci

Introduction

In recent decades, education systems have undergone profound transformations in response to the societal challenges of globalisation, including changes in labour markets, migration flows and individualisation processes (van de Werfhorst 2014). The effectiveness of traditional centralised models of state education in providing responses to the emerging educational needs of schools and families has been questioned. Inspired by economic theory (Chubb and Moe 1990) and recommendations from international organisations, such as the OECD and World Bank (Sellar and Lingard 2013), many countries have progressively adopted educational policies oriented towards the promotions of quasi-markets in education (Le Grand and Bartlett 1993; Ichilov, 2012). Proponents of neoliberal education policies (NLPs) propose educational reforms that aim to increase school autonomy (Levačić 1995; Clark 2009; Demas and Arcia 2015), promote competition between schools (Lubienski 2003; Musset 2012), incentivise forms of privatisation (Lieberman 1989; Urquiola 2016), and increase the accountability of principals and teachers (Bruns et al. 2011), while proposing standardised forms of student evaluation and assessment (Menashy, 2007).

However, many academics and left-wing observers have criticised neoliberal policies as morally questionable (Ball 2012) because they promote school segregation (Valenzuela et al. 2014) and lead to an increase in inequality of educational opportunity (Connell 1993; Whitty 2002). Building on this debate, a democratic discourse in education has re-emerged in recent years, proposing a wide range of democratic inclusive policies (DIPs) to promote equality, inclusion and social justice (Perry 2009). Although there is a wide variety of approaches and principles to democratic education (Sant, 2017), proponents of DIPs share the goal of developing a collaborative effort involving both school staff, beneficiaries, and surrounding actors to build cohesive educational communities (Sanders 2001; Epstein 2018).

Despite the importance of the debate from both an academic and a policy perspective, the empirical investigations on the diffusion and consequences of NLPs and DIPs is quite limited. Existing empirical studies mostly focus on historical overviews of educational reforms (Mundy et al., 2016), evaluation of single-policies (Blimpo et al. 2011), and qualitative case studies (Vulliamy 2004; Ball 2012), while evidence on large cross-national studies is more limited (Dumay & Dupriez 2014). This paper aims to fill this gap by providing a first descriptive overview of the distribution of school-level NLPs and DIPs across countries and by assessing the consequences of different types of policies on students' educational achievement and related inequalities. The focus on the school level allows us to gain a fine-grained understanding of how general policy recommendations have been applied and translated into educational practice, while also allowing us to assess the heterogeneity of policy applications within countries. The research strategy is divided into four main steps. First, we identify a set of key dimensions distinctive of NLPs and DIPs. Second, we provide a classification of policies based on the highlighted dimensions. Third, we map the distribution of policy profiles across the countries studied. Finally, we assess the relationship between policy profiles and students' educational achievement and inequalities in both language and mathematics competencies.

Data, variables, and methods

We use data from the OECD Programme for International Student Assessment (PISA) survey in 2022. PISA collects a wide range of information on educational performance and practices at different levels, including student, teacher, parent and school. For our purposes, we use information from the school questionnaire administered to school principals. Our analytical sample includes 19,858 school distributed in 80 countries. Based on the review of the literature and on the items available on PISA 2022 school questionnaire, we identify 10 dimensions characterising school policies (including admission restrictions, standardized tests, teachers' evaluation, inclusion activities, etc.).

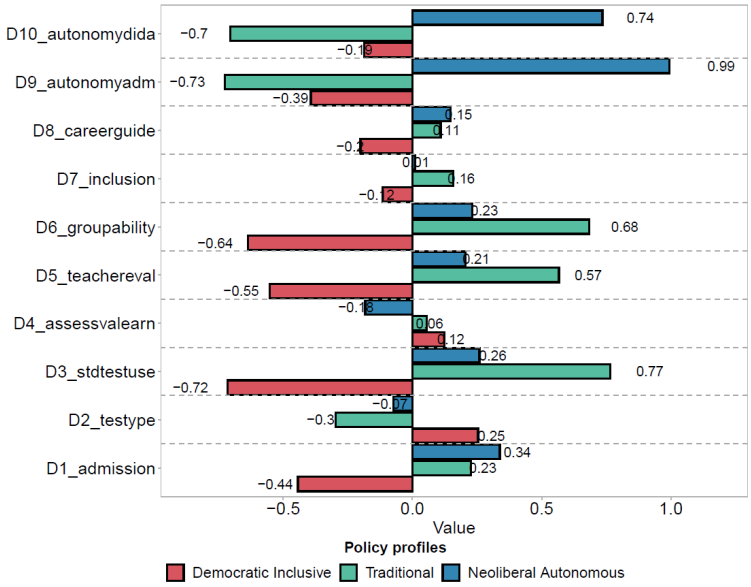
For each dimension, we compute a factorial index based on the batteries of items provided in the school questionnaire, using polychoric principal component analysis (PCA), as the items in the questionnaire are operationalised as nominal or ordinal variables (Kolenikov and Angeles, 2004). The indexes are expressed as cardinal variables and are standardised, with a mean equal to 0 and a standard deviation equal to 1. Next, we apply Latent Profile Analysis (LPA) to derive our school policy profiles (Spurk et al, 2020), and we compute the average probability of schools being classified in the policy profiles. To capture the multifaceted nature of school policy outcomes, we rely on various outcomes: 1) the school's average

mathematics test score and the proportion of students in the top 10% as measures of educational achievement and excellent performance, 2) the standard deviation, the share in the bottom 25%, and the SES-competencies association as measures of inequality in achievement. To investigate the relationship between policy profiles and a set of educational outcomes in mathematics we perform multilevel models with country fixed effects integrated with entropy balancing matching (EBM) (Hainmuller 2012). Country fixed-effects enable us to control for unobserved heterogeneity across countries relying on within-country estimator; the application of EBM with multivalue treatment allows us to control for a rich set of school characteristics that might confound the estimates of the effect of the school-policy profile (e.g. school principal features; school building characteristics and location; school's socio-demographics). In this abstract we report the bivariate association and the association conditional on the economic, social and cultural status of the students (ESCS index) to account for the association net of differences in the socio-economic composition of schools. In the next steps to be presented in the conference, we will account for differences in school-level characteristics and composition in a more systematic way.

Preliminary results

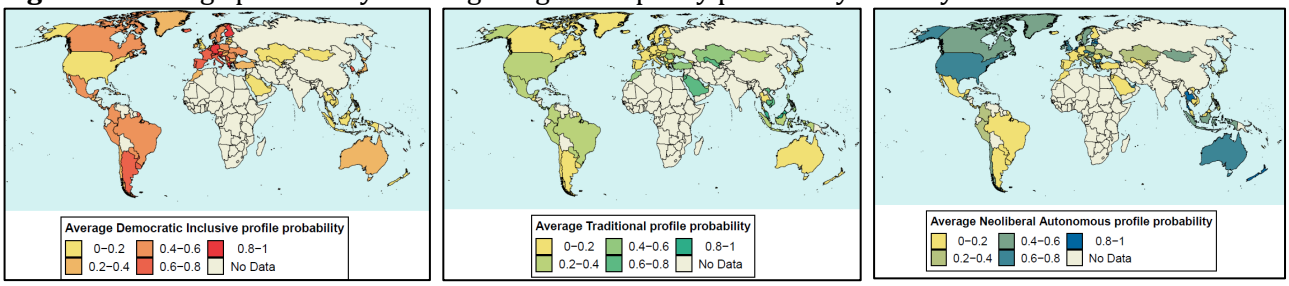
Preliminary results indicate three main policy profiles derived from LPA. The policy profiles, that we label as *Democratic Inclusive*, *Traditional* and *Neoliberal Autonomous* have different and distinctive features in respect of the 10 dimensions considered (Figure 1). Based on the distribution of factor loadings, the identification of policy profiles is grounded on two main axes: a) the content of school policies, and b) the degree of schools' autonomy. On the first, axis, the two Neoliberal profiles show positive factor loadings on a number of dimensions representing schools' propensity to impose restrictions on student admissions (D1), to use standardised tests (D3), to evaluate teachers' performance (D5), and to group students by ability (D6), while the Democratic Inclusive profile shows negative values on these dimensions. On the second axis (D9 and D10), the Neoliberal Centralised and Democratic Inclusive profiles show a lower degree of autonomy than the Neoliberal Decentralised profile. The distribution of policy profiles is quite balanced, with a 39.5% of schools assigned to the Democratic Inclusive profile, 34.5% to the Neoliberal Autonomous, and 26% to the Traditional profile.

Figure 1. Average values on dimensions used in LPA by Latent Profiles.



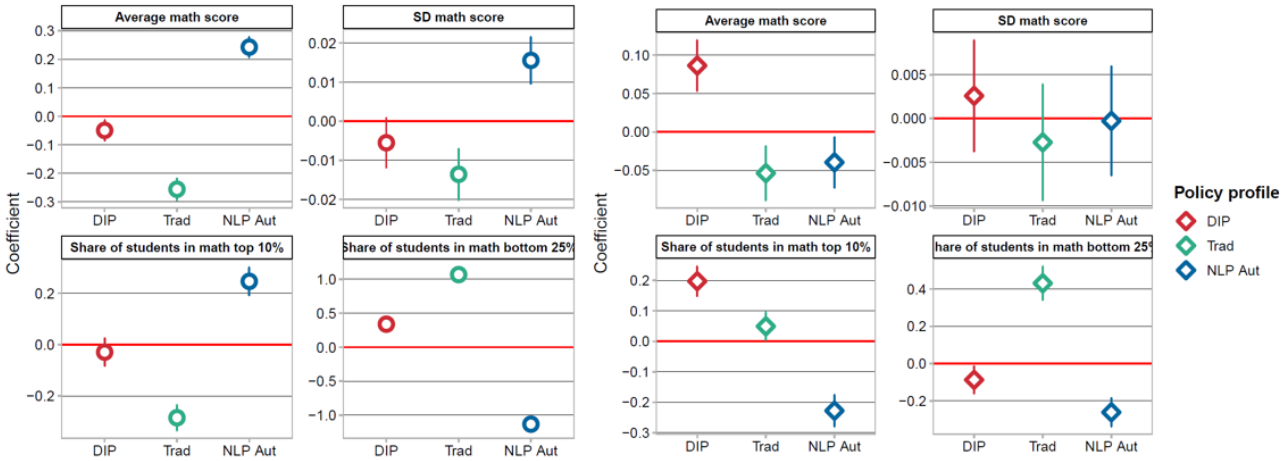
The distribution of the average probability of being assigned to the three profiles by countries follows a clear pattern (Figure 2). Schools in European and South American countries show a higher average probability of being classified in the Democratic Inclusive profile, schools in Middle East and South-East Asia show a higher probability of being classified in the Traditional profile, while schools identified in the Neoliberal Autonomous profiles are prevalent in Anglo-Saxon countries and some countries in Eastern Europe and South-East Asia. These findings are consistent with the literature, suggesting that the adoption of NLPs has been more prevalent in countries with liberal welfare regimes and in the emerging economies of Southeast Asia (Rizvi et al., 2006; Rutkowski, 2007).

Figure 2. Average probability of being assigned to policy profiles by country.



The fixed effects coefficients from the simple regression models (Figure 3) show that Neoliberal Autonomous schools have better results in mathematics, while also having a higher proportion of students in the top 10% and a lower proportion of students in the bottom 25%. However, the Neoliberal Decentralised schools have a higher dispersion, indicating greater inequalities in performance compared to the other profiles. The Traditional profile shows the worst performance in mathematics, while the Democratic Inclusive profile lies in between. However, controlling for school ESCS reveals that the positive effects on test scores and the share of high achievers displayed by Neoliberal Autonomous schools are largely explained by the socio-economic composition of students. Indeed, accounting for ESCS, schools with a Democratic Inclusive profile perform better in terms of average mathematics scores and proportion of high achievers compared to both the Neoliberal Autonomous and Traditional schools.

Overall, this study advances the debate on education policy and social inequality by revealing how policy orientations can reinforce or mitigate educational disparities, providing a robust empirical foundation for understanding the relationship between educational governance models and social stratification.



Keywords: Education policy, School governance, Achievement inequality, Latent profile analysis

Reducing the teaching penalty: the impact of an experience-based salary reform on second career teacher recruitment

Jacob Van Belle – Willem De Cort – Mike Smet

Introduction

Second career teachers (SCT) - individuals who enter the teaching profession at a later age after working in another sector - are considered a promising group to tackle teacher shortages across the globe (Haselkorn & Hammerness, 2008; Hogg et al., 2023; Siostrom et al., 2023; Tigchelaar et al., 2010). However, transitioning into teaching is often financially unattractive due to the so called "teaching penalty": teachers tend to earn less compared to other professions with comparable qualifications and responsibilities (Allegretto et al., 2008). This penalty is exacerbated for SCTs when their prior work experience is not recognized, resulting in entry-level salaries despite extensive professional backgrounds. Thus, one potential strategy to address this issue is the implementation of an experience-based salary policy that acknowledges relevant non-teaching experience, thereby narrowing the wage gap. However, no causal evidence exists on the effectiveness of such targeted salary policies (De Witte et al., 2023).

To fill this gap, we exploit the introduction of an experience-based salary policy in Flanders (i.e., the Northern, Dutch speaking part of Belgium) between September 2020 and September 2023. Before September 2020, SCT received the pay of a starting teacher without experience. However, a non-retroactive salary policy was introduced in September 2020 with a recognition of up to 8 years of professional experience outside of the teaching profession for new teachers for certain subjects with great shortages. This was further expanded to 10 years of experience and more subjects in December 2021, and finally to 15 years of experience for all teaching positions in September 2023 (Omzendbrief PERS/2020/04 van 15/07/2020, 2024). This natural policy experiment affected some groups of teachers (i.e., SCT with an eligible degree, who start after a certain date) but not others. Figure 1 graphically shows when and for what subjects the seniority policy was introduced. Based on our simulations with the publicly available salary simulator of the Flemish Department of Education & Training, we estimate that this policy can have a significant impact on the wage of a starting SCT with the current salaries, with salary increases between 5 and 25% depending on the degree and prior experience.

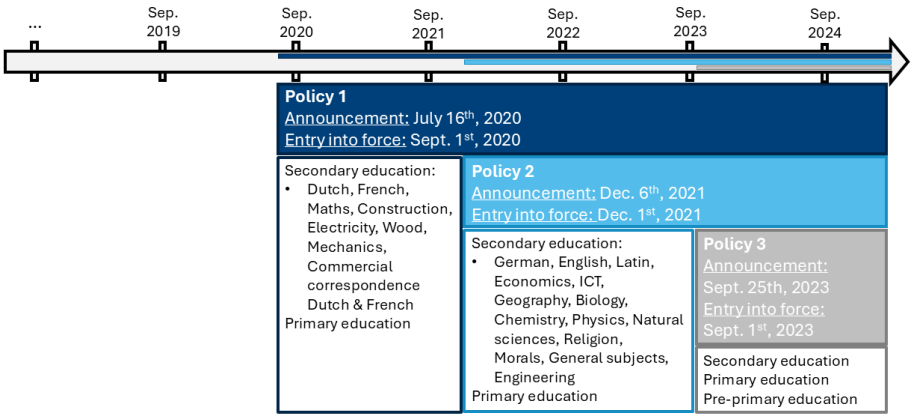


Figure 1: Overview of the staggered implementation of the experience-based salary policy

Research object & hypothesis

Our aim is to evaluate the effectiveness of the staggered introduction of an experience-based salary policy directed at SCT. We answer the following three research questions:

1. **Enrolment in teacher education:** Did the policy increase the share of older (≥26y) students enrolled in an eligible teacher education program increased?

2. **Job-entry of teachers:** Did the policy increase the share of older teachers (≥26y or ≥30y) with an eligible degree entering the teaching profession, increased?

3. **Transition from teacher education towards teaching profession:** Did the policy increase the transition rate from teacher education towards the profession increased among eligible older teachers (≥26y or ≥30y)?

We hypothesized a significant positive effect on the share of older teachers enrolled in an eligible teacher education program, on the probability that a starting teacher with an eligible degree is an older teacher, and on the probability that an eligible older teacher transitions from teacher education towards the teaching profession.

Data

We rely on three different administrative datasets to answer our research questions. The first dataset (research question 1) is a publicly available dataset with aggregated data at the level of a specific teacher education program in a higher education institution. It contains the number of students who started a teacher education program at a Flemish higher education institution in a given year and some background variables such as age group at the time of enrollment (<26y or ≥26y). Due to data restrictions, we were unable to use observations prior to 2019-2020 or from bachelor's degrees for secondary education. Due to all subjects being treated from 2023-2024 onwards, we also omitted that school year. Our final dataset has 278 observations in 70 different teacher education programs in 21 different higher education institutions between 2019-2020 and 2022-2023. The dependent variable is the share of students aged 26 years or older in a specific teacher education program in a specific academic year. We reason that eligible teachers, who have built up seniority that can be recognized, are all part of the group aged 26 years or older. Some first career teachers can also be part of this group, meaning that we can have measurement error. However, this can only lead to a lower bound estimate of the actual treatment effect.

The second dataset (research question 2) contains information on all teachers and their separate assignments in the Flemish pre-primary and compulsory education during the school years 2018-2019, 2020-2021, and 2022-2023. From those teachers, we retained all starting teachers in one of those three school years during their first year in education. We have information on all higher education degrees a teacher obtained (both teacher education degrees and other degrees), and other background information such as age or sex. We also have detailed information on assignment characteristics, but we cannot uncover the exact subject(s) a teacher taught. Therefore, to identify eligible teachers, we rely on the higher education degree(s) one obtained. Again, we have some measurement error. At worst, we will overestimate the number of eligible teachers as it can be that one has an eligible teacher degree but taught an ineligible subject. Our final dataset has 17,302 starting teachers. The dependent variable is the probability that a starting teacher in a year is an older teacher. We consider two specifications: (1) in line with RQ1, we consider an older teacher to be ≥ 26y, or (2) we consider an older teacher to be ≥ 30y. Note that in the second operationalization, we consider younger teachers still to be <26y, omitting teachers between 26y and 30y old.

The third dataset (research question 3) contains all teacher education graduates in Flanders between the school years 2008-2009 and 2021-2022 and their teaching assignments, if any, until June 30th, 2022. In this dataset, we only observe people that graduated from teacher education. We have information on the exact teaching degree one obtained, some background variables such as age or sex, the exact start and end date of every separate teaching assignment and detailed assignment characteristics of every assignment. This allows us to identify whether a starting teacher was eligible for the recognition of their seniority or not. We operationalize old teachers in the same way as for the second dataset. Our dependent variable is the probability that a teacher education graduate enters the teaching profession within 1 year after graduation. Identification is again not perfect, but measurement error will most likely only create an underestimation of the true treatment effect. Our dataset contains 67,143 graduate teachers. However, we focus on teachers graduating in 2017 or later, resulting in 23,525 graduates.

Method

To exploit the natural experiment and to estimate causal effects, we rely on difference-in-differences methods, integrating novel developments in this area (Baker et al., 2025). To answer the first two research questions, we employed the difference-in-difference (time & degree) with multiple periods estimator that was proposed by Callaway & Sant'Anna (2021) via the csdid package in Stata (Rios-Avila et al., 2021). This estimator allows us to study the average treatment effect, but also the treatment effect in each group and how that evolves over time. To answer the third research question, we leverage a triple difference situation (time & degree & age-group) to estimate the probability that an eligible teacher transitions from teacher education towards the teaching profession within a year upon graduating, manually implementing the estimators proposed by Ortiz-Villavicencio & Sant'Anna (2025), as to date no code has been released to that

end. We implement 2 separate csdid specifications (time & degree), comparing older teachers with younger teachers, without specifying additional covariates.
In the full paper, we will additionally conduct robustness analyses (i.e. conducting placebo tests by altering the treatment time(s), verifying the assumption of no compositional changes), study heterogeneous effects (e.g., based on the education level of the teacher degree), and conduct a cost-effectiveness analysis of the policy.

Results & findings

Our main findings on the effectiveness of the policy present a mixed picture. First, we observe a significant and substantial overall increase of the share of older students (≥26y) enrolling in the treated teacher education programs with 5.39% (p=0.012). Splitting this out by treatment group, there is an increase of 4.94% (p = 0.047) in the group of teacher education programs treated since 2020 and of 6.78% (p = 0.032) in the group treated in 2021. Figure 2 presents the results.

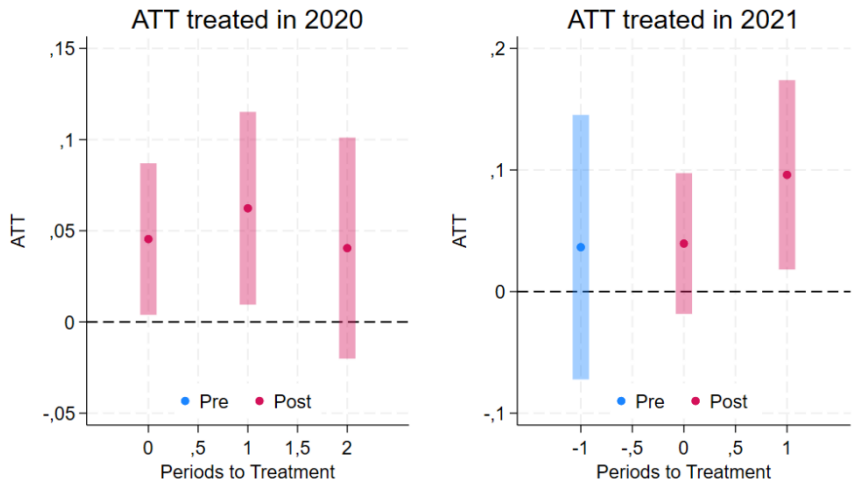


Figure 2: Treatment effect on the enrolment in eligible teacher education programs (RQ1).

Second, there is a slight but insignificant decrease on job-entry among older teachers, regardless of the treatment group (2020 or 2021) or the operationalization of older teachers (≥26y or ≥30y). Looking at the probability that a starting teacher is at least 26 years old, than the ATT in treatment group 2020 is -0.0147 (p = 0.371) and -0.0076 (p = 0.762) in the treatment group 2021. Defining older teachers as being at least 30 years old, than the ATT in treatment group 2020 becomes -0.0126 (p = 0.405) and in treatment group 2021 it is -0.0094 (p = 0.701). The treatment effects are presented in figure 3.

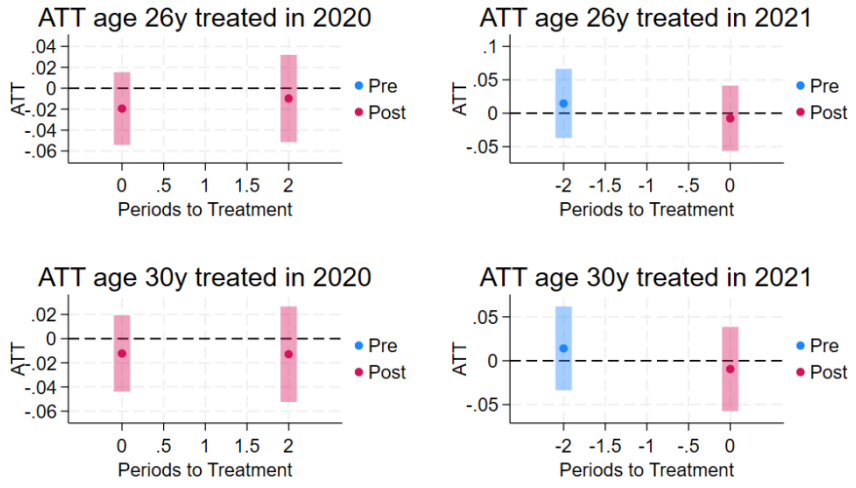


Figure 3: Treatment effect on the immediate job-entry of eligible teachers (RQ2).

Third and last, there is an insignificant increase in the transition rate among older teachers in both considered specifications (≥26y & ≥30) and in both treatment groups (2020 and 2021). Among teachers ≥26y, the ATT of the group treated in 2020 is 0.0195 (95%CI[-0.0302, 0.0692]) and that of the group treated in 2021 is 0.0078 (95%CI[-0.0600,0.0756]). If we only consider teachers of 30y or older, then the ATT for the 2020 group is 0.0084 (95%CI[-0.0458,0.0626]) and for the 2021 group 0.0442 (95%CI[-0.0201,0.1085]). The treatment effect is clearly large in the group that was treated in 2021 when we only consider teachers ≥30y. Figure 4 shows these estimated treatment effects graphically.

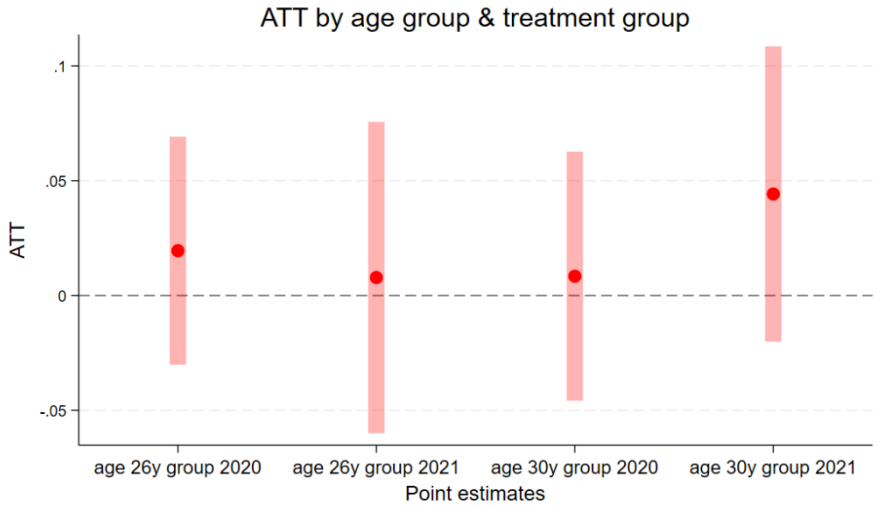


Figure 4: Treatment effect on the transition between teacher education and the teaching profession (RQ3).

To conclude, our main findings show a significant and sizable increase on the enrolment of SCT in eligible teacher education programs, but not on other outcomes. Hence, the short-term effectiveness of the salary reform is limited. This difference in effectiveness by outcome variable has two explanations. First, we expect that most people already enrolled in a teacher education program have decided on joining the profession or not beforehand, and their decision will not greatly change due to the announcement of a salary policy. Second, it is likely that positive effects on job-entry and on the transition from higher education towards the teaching profession are not immediate, given that those convinced by the policy still need to finish their teacher education before entering the profession. This might take multiple years as the study duration of a teaching degree ranges between 1 and 3 years for a student in a standard learning track.

Keywords: Teacher shortages, Second career teachers, salary policy, difference-in-differences

SESSION 3. INTERNATIONAL LARGE-SCALE ASSESSMENTS (ILSAS) METHODS AND RESULTS

ORGANIZER: UNIVERSITY OF BATH
COORDINATOR: ELLEN CLAES
21ST NOVEMBER: 8.30 A.M. - 10.45 A.M. {AULA 4 – RESEARCH 13}

Children of Immigrants' Academic Effort. Evidence from PISA computer-based tests in 56 countries

Eleonora Vlach – Loris Vergolini

Introduction and theoretical background

This study aims at understanding if children of immigrants show greater academic effort in comparison to children of natives, and if institutional features such as the selectivity and stratification of the educational system influence children of immigrants' academic effort.

Previous empirical studies converge in reporting a generalised gap in educational achievements for children of immigrants with respect to natives, both across countries and across ethnic minorities (Marks, 2005; Heath & Brinbaum, 2014; Drouhot & Nee, 2019). However, this generalised gap in achievements does not seem to translate into an identical gap in educational attainment (Jonsson et al., 2011, Jackson, 2013). The term "ethnic paradox" has been thus suggested with reference to this phenomenon (Hill & Torres, 2010; Becker & Gresch, 2016). When it comes to educational decisions, many children of immigrants outperform natives with the same level of socio-economic resources and school grades. The most frequently suggested reason for this is the higher educational ambitions characterising children of immigrants (Goyette & Xie, 1999; Fernández-Reino, 2016; Rudolphi & Salikutluk, 2021; Dollmann et al., 2023).

The present study seeks to contribute to the debate by shedding new light on the "ethnic paradox". We do so by focusing on a dimension largely unexplored so far: that of non-cognitive skills. To the best of our knowledge, our study is the first one examining non-cognitive skills of children of immigrants in comparative perspective. We aim also at contributing to the debate on the role of macro-societal features in the integration processes (van Tubergen et al., 2004; Levels et al, 2008; van de Werfhorst and Heath, 2019). Therefore, we compare children of immigrants' academic efforts in 56 countries worldwide, in order to understand which structure of opportunity (here: the educational system) more effectively minimises the intergenerational transmission of the ethnic disadvantage through education.

The idea that despite lower grades and scholastic failure, children of immigrants maintain high educational ambition is intriguing, due to the high potential for long-term integration it represents. However, the assumption that children of immigrants' aspirations do effectively translate into pro-school behaviours, has remained largely untested so far. In line with the literature on immigrants' aspirations, we first test if *children of immigrants show comparatively higher academic effort than their peers independently from the country of destination* (H1).

From a theoretical point of view, the explanatory power of theories on immigrants' aspirations is not context specific. This reasoning leads us to our second hypothesis (H2a) that *there is no cross-country variation in the higher level of academic effort of children of immigrants.* However, previous studies in sociology of education suggest that the education system does play a role. Jackson & Johnson (2012), for instance, drew a separation line between comprehensive systems where educational progress is mainly a matter of individuals' preferences, and stratified/selective systems where progress through levels is only allowed against the successful performing of competitive exams. Hence, in contrast to H2a, a hypothesis can be put forward that: *students' migration background has a greater positive effect on academic effort in choice-based- in comparison to exam-based educational systems* (H2b).

Data and Analytical Strategy

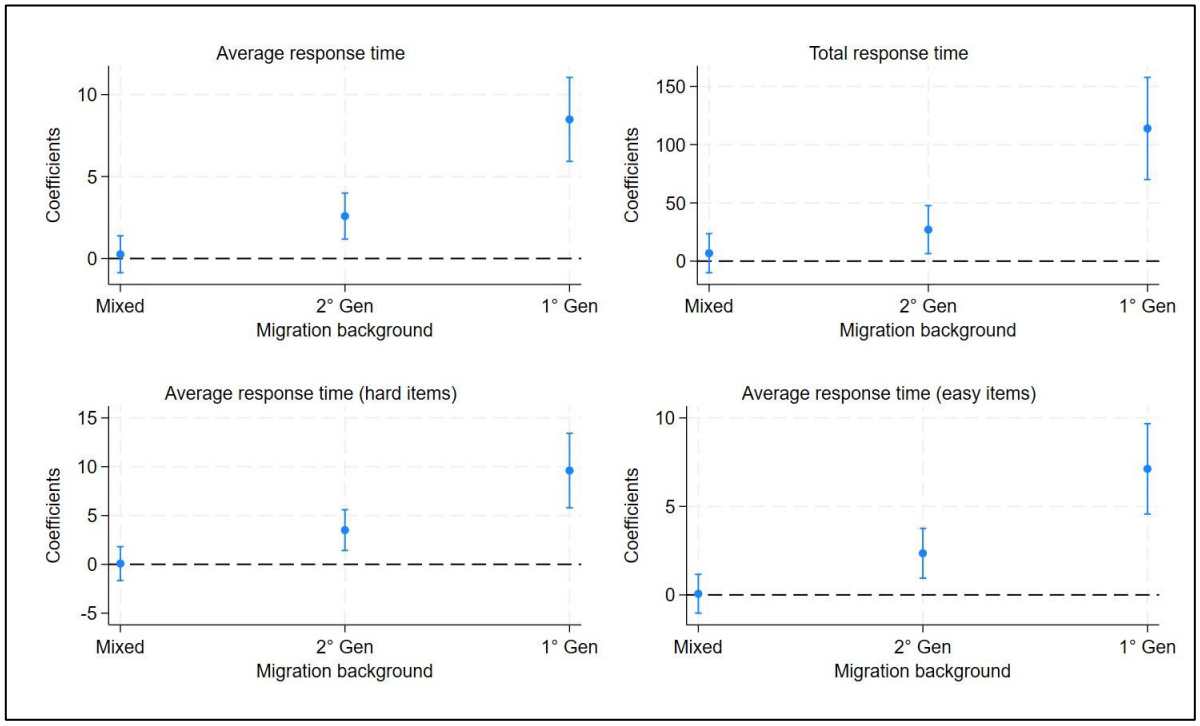
To test our hypotheses, we employ data from PISA 2015: the first computer-based PISA survey. Computer-based assessment allows the production of a set of digital traces (the so-called log-files) that can be exploited for educational research. This data supplies a very detailed set of information on how students specifically responded to the test, covering a vast set of information such as: response time, number of actions/clicks, and browsing. In line with the related field literature, we measure students' effort as response time (Wise & Kong, 2005). To assure the robustness of our conclusions, we consider four different

measures: the average response time, the total response time, the average response time to the most difficult items, and that of the easiest ones.[1]
We only analyse answers to the first cluster of the PISA test.[5] Our main analyses are carried out through OLS regressions having migrant generation as the main explanatory factor. Models control for a rich set of covariates, including gender, age, social origins, number of books at home, familiarity with ICT, language spoken at home, PISA reading score, and school fixed effects. Besides the "usual suspects" (i.e., social origins and socio-demographic characteristics), we also include language spoken at home and the student's score at the reading test to make sure longer response time is not due to differences in literacy. The OECD index of familiarity with ICT is included to control for potential differences in ability to respond to computer-based stimuli.

Results

In line with the theoretical expectations, our results suggest that children of immigrants, show more academic effort than natives in performing the test. This finding is mostly driven by first generations, with a statistically significant difference between first and second generations (Figure 1).

Figure 1. Estimated effect of the migration generation on academic effort in terms response time (expressed in seconds). OLS regressions. Weighted results.



To scrutinise the role played by institutional context, we estimate a set of separate models according to the degree of selectivity of the educational system[3] and the macro geographical area (Figure 2). Looking at the first row of Figure 2 reporting on the Western Europe, there emerge no differences across education systems in immigrants' greater effort. This result goes in the direction of H2a. However, the picture changes when we focus the attention on other geographical areas: other OECD countries (second row) and non-OECD countries (third row). Outside the European Union, the hypothesis receiving higher support seems to be H2b.

In order to better understand the differences in academic effort that may depend on the presence of tracking, we additionally predicted the value of academic effort in a model interacting migrant background and upper secondary school track (Figure 3). All students enrolled in an academic track show higher

[1] Item difficulty is supplied by OECD and is calculated through Rash model (OECD, 2017). Here we look at the five most difficult items and at the five easiest ones.

[2] The PISA test is divided into four clusters. The rationale for analysing only the first is to avoid any potential bias induced by the onset of fatigue, which increases response time disproportionally across students in the later clusters.

[3] The degree of selectivity is built considering jointly the indicators of horizontal and vertical differentiation supplied by OECD (2010).

response time than those enrolled in a vocational track, whereas the immigrant premium in academic effort results particularly strong in the academic track. Rather than the formal setting of the education system, it thus seems that the specific situation children of immigrants more closely experience in their school life (i.e. knowing that the school they are attending will allow them to remain longer in full time education) is the one conditioning their behaviours the most.

Figure 2. Estimated effect of the migration generation on academic effort in terms response time (expressed in seconds) according to the differentiation of educational system and geographical area. OLS regressions. Weighted results.

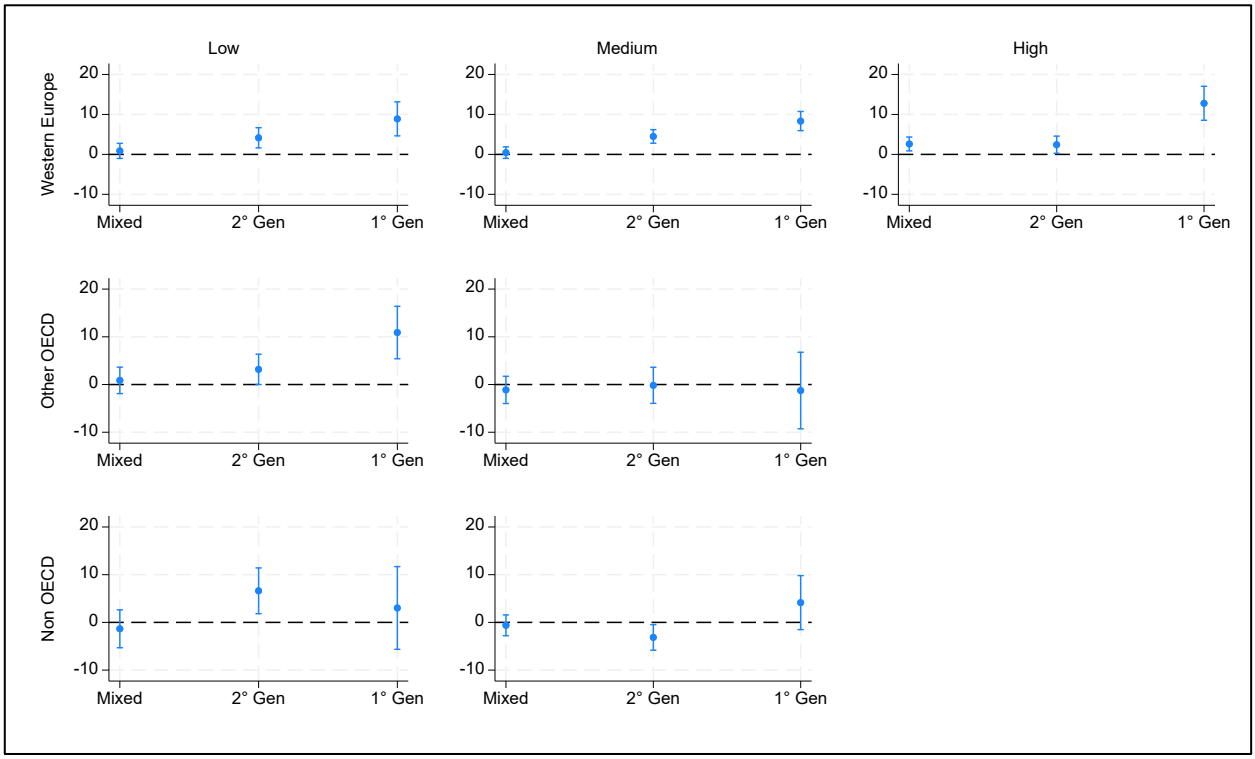
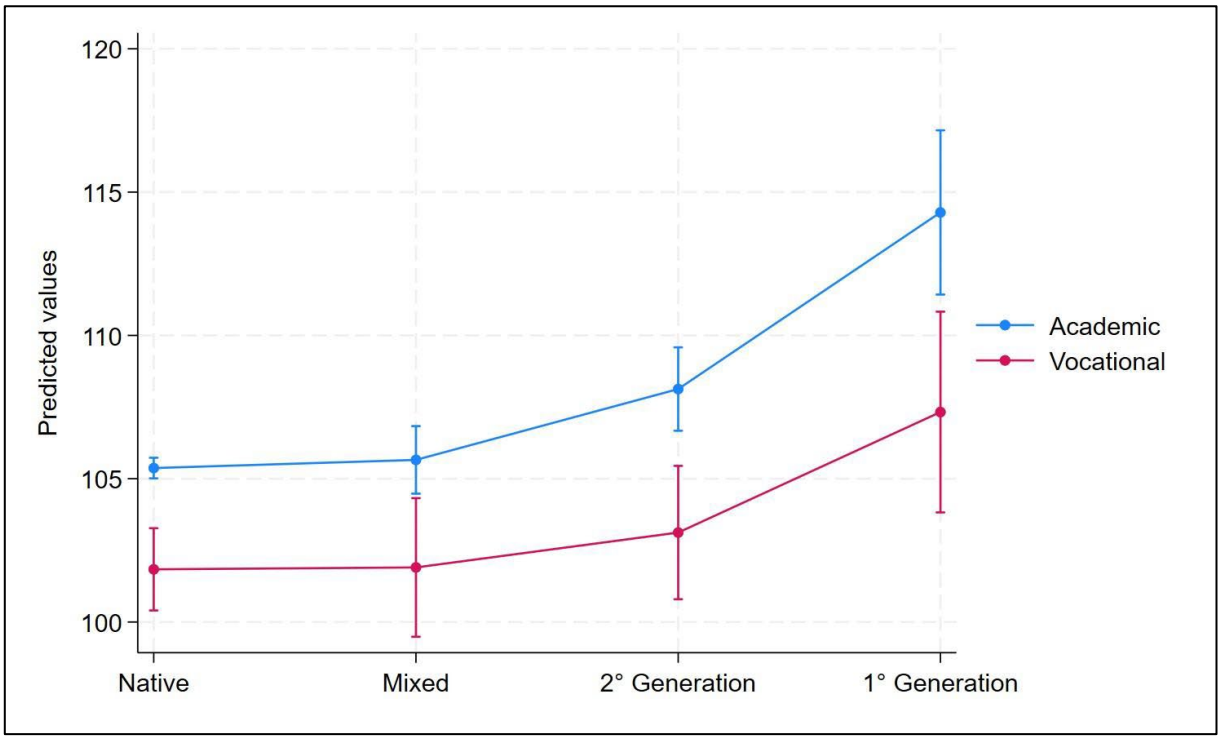


Figure 3. Predicted values of the academic effort in terms response time (expressed in seconds) according to migration generation and high school track. OLS regressions. Weighted results.



Keywords: Ethnic inequality, non-cognitive skills, educational system, PISA data

Measuring Digital Self-Efficacy in International Large-Scale Assessments: An International Comparison Between ICILS and PISA

Juan Carlos Castillo – Daniel Miranda – Tomás Urzúa

Introduction

Digital self-efficacy (hereinafter DSE), conceived as expectations about one's capabilities to learn and accomplish tasks in digital technologies and digital environments, is one of the principal components to promote the formation of digital competences (Ulfert-Blank and Schmidt 2022). DSE is a construct frequently measured in international large-scale assessments (ILSAs), as substantial evidence indicates its critical role as an explanatory variable in the development of digital competences within educational settings (Scherer and Siddiq 2019; Hatlevik et al. 2018; Claro et al. 2012). Furthermore, studies consistently demonstrate that DSE also allows individuals to acquire and apply digital skills effectively along the life spam (Rohatgi, Scherer, and Hatlevik 2016; Scherer, Siddiq, and Teo 2015).

The conceptualization and operationalization of DSE vary notably in terms of concepts and their measurement. Some studies conceive DSE as an unidimensional construct, measuring individuals' overall confidence in using digital technologies without distinguishing between types of tools and/or levels of complexity (Hatlevik, Guðmundsdóttir, and Loi 2015; Rohatgi, Scherer, and Hatlevik 2016). In contrast, other studies adopt a multidimensional approach, distinguishing between general and specialized self-efficacy to account for the nature and complexity of digital tasks (Scherer, Siddiq, and Teo 2015). Whereas *general DSE* encompasses confidence in everyday tasks such as internet navigation or word processing, *specialized DSE* involves more advanced activities such as programming and/or data analysis (Ulfert-Blank and Schmidt 2022). The distinction between general and specialized DSE is particularly relevant in the context of ILSAs, where the measurement of digital self-efficacy can significantly influence the interpretation of students' digital competencies. For instance, a unidimensional measure may aggregate all digital task-related confidence into a single score, potentially masking important differences in how students perceive their abilities across different digital contexts. In contrast, a multidimensional measure allows for a more precise understanding of students' digital competencies and can inform educational policies and practices more effectively.

Between the two most relevant ILSAs in the Digital Competence agenda (ICILS and PISA), a critical inconsistency persists in their conceptualization and measurement of DSE. PISA operationalizes DSE as an unidimensional construct, aggregating all digital task-related confidence into a single generalized measure (OECD 2021). In contrast, ICILS adopts a bidimensional framework, distinguishing between general DSE (basic digital tasks) and specialized DSE (advanced tasks) (Fraillon et al. 2020; Scherer and Siddiq 2019). This discrepancy raises essential questions about construct validity and cross-assessment comparability, particularly since the choice of model (unidimensional vs. multidimensional) may influence policy interpretations and pedagogical interventions. For instance, unidimensional models could underestimate the predictive power of DSE for complex digital problem-solving, as it is shown in studies that specialized DSE is a stronger predictor of performance in advanced digital tasks than general DSE (Scherer and Siddiq 2019). Besides, some studies show that while gender gaps in general DSE are minimal or non-existent, women tend to report significantly lower confidence in specialized DSE domains—particularly those involving STEM-related digital tasks (Hargittai and Shafer 2006; Cai, Fan, and Du 2017; OECD 2021). Therefore, understanding the proper use of the dimensions of DSE in two of the most important ILSAs including DSE measurement is necessary to refine the scientific use of this construct and to understand different populations' expectations with technologies.

The present study's aim is (i) To assess the bi-dimensional approach of DSE in International Large-scale Assessment studies, and (ii) To evaluate the comparability of the bi-dimensional measurement of DSE across countries and gender in ICILS and PISA. To achieve this, we will conduct confirmatory factor analysis (CFA) and measurement invariance testing using data from the latest cycles of ICILS (2023) and PISA (2022). This approach will allow us to rigorously evaluate the validity of the two-dimensional model of DSE and its cross-cultural applicability. CFA is particularly well-suited for testing theoretical models where specific latent structures are hypothesized a priori—such as the proposed distinction between general and specialized dimensions of digital self-efficacy. Besides, measurement invariance testing is crucial for ensuring that the same construct is being measured across different groups and contexts, which is essential for making valid comparisons of DSE across countries.

1. Research Questions and Hypotheses

The present study aims to address the following research questions:

1. Is it possible to identify two latent dimensions of digital self-efficacy (general and specialized) based on related batteries and indicators, in a comparable way between PISA and ICILS?
2. Is the bi-dimensional measurement model of digital self-efficacy equivalent between girls and boys?
3. Is the bi-dimensional measurement model of digital self-efficacy equivalent across countries?

To answer these questions, we will test the following hypotheses:

H1: It is possible to identify two latent dimensions of digital self-efficacy (general and specialized) based on related batteries and indicators included in large-scale assessments such as PISA and ICILS (bi-dimensional hypothesis)

H2: The bi-dimensional measurement model of digital self-efficacy is invariant (in terms of measurement) girls and boys.

H3: The bi-dimensional measurement model of digital self-efficacy is invariant across countries.

Methods

Data

We have two main data sources. The first one is ICILS, developed by the International Association for the Evaluation of Educational Achievement (IEA). We use data from the third wave (2023), which encompasses 35 educational systems, testing 135.615 8th grade students on computer and information literacy (CIL) and computational thinking. The study evaluates students' ability to use digital tools responsibly, problem solving, and collaborate online. Data is collected through performance tests and contextual questionnaires for students, teachers, and schools. A key feature of ICILS is its bidimensional measurement of digital self-efficacy (DSE), distinguishing between general and specialized.

The second data source is PISA, implemented by the OECD. It assesses 15-year-olds' skills in mathematics, science, and reading across multiple cycles (the last three ones in 2015, 2018, 2022). The study's primary objective is to evaluate education systems' effectiveness in preparing students for future challenges, with a growing emphasis on digital readiness. The 2022 assessment covered 81 countries/economies with a

sample exceeding 600,000 students. Digital self-efficacy (DSE) was last measured in 2022 as part of the optional ICT familiarity questionnaire, following its absence in the 2018 cycle. This questionnaire was applied in an optional way in 53 countries, which are included in the analysis (N = 279,435).
Both data sources are publicly available and can be accessed through the IEA and OECD websites, respectively. ICILS 2023 data can be found at https://www.iea.nl/studies/iea/icils/2023, whereas PISA 2022 data is available at https://www.oecd.org/pisa/data/pisa-2022-database/.

Variables
The analysis will focus on the digital self-efficacy (DSE) items from both ICILS and PISA. The DSE items in ICILS 2023 are designed to measure students' confidence in performing various digital tasks, while PISA 2022 includes similar items but framed within a unidimensional context.
Table 1 summarizes the measurement batteries for self-efficacy in both studies in a comparative way:
Table 1: ICLS and PISA items comparison

Task Category	**ICILS 2023 Item**	**PISA 2022 Item**
Search information	Search for and find relevant info for a school project	Search for and find relevant information online
Assess information	Judge whether you can trust information you find	Assess the quality of information you found online
Create multimedia	Create a multi-media presentation	Create a multimedia presentation
Edit documents	Insert an image / edit text for assignment	Write or edit text for a school assignment
Edit images	Edit digital photographs	—
Upload/share content	Upload or share content	Share practical information / explain sharing
Collaborate	Collaborate on group assignment	Collaborate with students
Change settings	Change device settings	Change settings to protect data/privacy
Install/select apps	Install programs	Select most efficient app
Programming	Write program in code	Create visual/text-based program
Build a webpage	Build/edit webpage	Create/maintain webpage or blog
Identify software errors	–	Identify source of a software error

Estimation methods
Statistical analyses will address missing data using Full Information Maximum Likelihood (FIML) during confirmatory analyses. Before modeling, PISA responses of "I don't know what this is" will be recoded as missing, and the distribution of missing values will be examined by country. For each country, a two-factor confirmatory factor analysis (CFA) will be conducted separately for PISA and ICILS data, distinguishing between general and specialized digital tasks. Model fit will be evaluated using chi-square, Comparative Fit Index (CFI), Tucker-Lewis Index (TLI), and Root Mean Square Error of Approximation (RMSEA) (Brown 2015).
Measurement invariance across gender and countries will be tested using multi-group CFA, progressing through configural, metric, and scalar invariance. Changes in model fit will be interpreted using established thresholds (e.g., $\Delta CFI \geq -0.004$; $\Delta RMSEA \leq .05$ for metric, and $\Delta CFI \geq -0.004$; $\Delta RMSEA \leq .01$ for scalar invariance) (Rutkowski and Svetina 2017). All items will be treated as ordered variables, and FIML will be used for missing data throughout (Enders and Bandalos 2001). Analyses will be conducted using the Lavaan library for R; the scripts and data are available at https://github.com/milenio-nudos/ILSAs_batteries_measurement.

Preliminary analyses and results

For ICILS 2023, a confirmatory factor analysis (CFA) estimation confirms the bi-dimensional model, with indices indicating good model fit (CFI = 0.97, TLI = 0.96, RMSEA = 0.045). In the case of PISA, the CFA results for the two-factor model show acceptable fit (CFI = 0.95, TLI = 0.94, RMSEA = 0.052), supporting the applicability of the bi-dimensional DSE model in both assessments, though with stronger evidence in ICILS. Regarding measurement invariance, the analyses indicate that for gender, both ICILS and PISA achieve configural and metric invariance, suggesting that the bi-dimensional model of digital self-efficacy is structurally similar and factor loadings are equivalent across girls and boys. However, scalar invariance is only partially supported, as some items exhibit differential item functioning by gender. In terms of country-level invariance, configural invariance is supported in both datasets, and metric invariance is achieved in most countries. Nonetheless, scalar invariance is more limited, particularly in PISA, indicating that some item intercepts vary across countries and may affect the comparability of mean scores.

Analysis of gender differences reveals that general digital self-efficacy (DSE) shows minimal disparities between girls and boys in both ICILS and PISA, while specialized DSE consistently favors boys, with effect sizes (Cohen's d) ranging from 0.20 to 0.35 across countries. Cross-country comparisons indicate that Nordic and East Asian countries tend to have higher average DSE scores in ICILS. Similar patterns are observed in PISA, although specialized DSE exhibits greater variability across countries.

Summary

- The bi-dimensional model of DSE is empirically supported in both ICILS and PISA, though with stronger evidence in ICILS.
- Measurement invariance is generally supported at the configural and metric levels, but scalar invariance is more challenging, especially across countries.
- Gender gaps are more pronounced in specialized DSE.

Keywords: Digital self-efficacy, ILSAs, confirmatory factor analysis, measurement invariance

Moving Beyond Averages: Climate Strength and the Classroom Experience in TIMSS 2019 and 2023

Zhijun Chen

Introduction

Teaching quality in the classroom plays a fundamental role in shaping student outcomes and addressing persistent educational inequalities (e.g. Liu et al., 2024; Teig et al., 2022). Research over the past two decades has converged on several key dimensions of teaching effectiveness that explain variation in student learning (Klieme et al., 2001; Lipowsky et al., 2009). One of the most widely accepted frameworks, particularly within the German-speaking education research community, conceptualizes teaching quality across three core dimensions: *classroom organization* (the ability to manage time, space, and behavior), *teacher support* (the emotional and motivational responsiveness of the teacher), and *cognitive activation* (the extent to which instruction challenges students intellectually) (Klieme et al., 2009). More recently, a fourth dimension, instructional clarity, has gained prominence as a vital indicator of teaching effectiveness, particularly in large-scale international assessments (Teig & Luoto, 2024). Instructional clarity refers to the teacher's ability to communicate goals clearly, provide understandable explanations, and structure lessons coherently.

Despite the theoretical sophistication of this model, large-scale studies often reduce teaching quality to class- or school-level averages of these dimensions, neglecting the extent to which students within the same classroom experience teaching practices consistently (e.g. Teig et al., 2022). This consistency, or lack thereof, can influence how learning opportunities are distributed among students. To advance the field, we incorporate the concept of climate strength, originally developed in organizational psychology, which captures the *within-class agreement* in students' perceptions of teaching right (Keeler et al., 2023; Schneider et al., 2002). While climate strength has been applied to study leadership effectiveness, workplace culture, and team cohesion, it remains underexplored in education research right (Olson & Jiang, 2021).. This study positions climate strength as a complementary lens that can reveal how shared classroom experiences amplify, or in some cases weaken, the impact of teaching practices on student achievement.

Research Object and Hypothesis

We examine the interaction between average levels of teaching practices and climate strength, focusing on two specific instructional dimensions: instructional clarity and classroom discipline. We hypothesize that these two practices are positively associated with student achievement, and that their effects are significantly stronger in classrooms where students share more consistent perceptions (i.e., stronger climates). Additionally, we explore whether these patterns have shifted between the 2019 and 2023 cycles of the *Trends in International Mathematics and Science Study (TIMSS)*, in light of broader changes in teaching conditions following the COVID-19 pandemic.

Country Selection Rationale

To contextualize the findings and enhance cross-national relevance, we focus on Italy as the main case and compare it with a set of Western European peers—France, Spain, Portugal, Ireland, and Belgium (both Flemish and French communities). These countries were selected for their similar educational governance structures, shared EU policy commitments, cultural proximity, and data availability in both 2019 and 2023 TIMSS cycles.

Data

Our analysis draws on student-level data from the Trends in International Mathematics and Science Study (TIMSS) 2019 and 2023 Grade 8 assessments, focusing on mathematics and science (Davier et al., 2024; Martin et al., 2020). Among the four dimensions of teaching quality, we selected instructional clarity and classroom discipline due to their theoretical relevance and consistent availability across cycles.

Mathematics Achievement

Student performance in mathematics is a core outcome of interest. TIMSS provides five plausible values of standardized mathematics achievement scores that reflect students' mastery of number, algebra, geometry, and data. These scores are internationally scaled and comparable across cycles and countries.

Science Achievement

We also examine science achievement with five plausible values, as TIMSS assesses students' understanding of biology, chemistry, physics, and earth sciences. This allows us to test whether the moderating role of climate strength applies consistently in science.

Teaching Practices and Climate Strength

Instructional clarity and classroom discipline are derived from student questionnaire responses. For each practice, we calculate the class-level mean (representing the average experience) and the within-class standard deviation (representing climate strength). A lower standard deviation indicates greater consensus among students and a stronger classroom climate.

Control Variables

To isolate the effects of teaching practices, we control for a set of background variables. These include student gender, socioeconomic status (SES) based on home possessions and parental education, and language spoken at home.

Method

We apply multilevel linear modelling (students nested within classrooms) to examine how instructional clarity and classroom discipline relate to student achievement, and how these relationships are moderated by climate strength. We include interaction terms between the mean and the standard deviation of each instructional variable to assess whether stronger climates strengthen the effects of teaching practices. Models are estimated separately for mathematics and science, and for each TIMSS cycle. Student-level controls such as gender and socioeconomic status (SES) are added to adjust for background disparities.

Contribution and Relevance

This study contributes to the ILSA literature in three key ways. First, it introduces climate strength as a theoretically grounded and empirically measurable classroom feature, offering added depth to traditional analyses of TIMSS data. Second, it demonstrates how student questionnaire data, often underused in cognitive outcome studies, can yield meaningful insights about instructional coherence. Third, by focusing on Italy and comparable European systems, the study provides policy-relevant findings at both national and regional levels, highlighting the importance of shared classroom experiences in enhancing teaching effectiveness. The methodological approach also serves as a proof of concept for expanding the use of ILSAs to explore more nuanced aspects of educational quality beyond simple mean comparisons.

Conclusion

This study contributes to the literature on teaching effectiveness by emphasizing not only what students experience in the classroom but how consistently they experience it. By integrating climate strength into the analysis of large-scale assessment data, we offer a deeper understanding of the classroom dynamics that support learning. For Italy and its European peers, improving teaching quality should involve efforts to

enhance both instructional delivery and coherence, ensuring that all students in a classroom benefit from high-quality teaching in a shared, structured environment.

Keywords: International large-scale assessments, teaching quality, climate strength, comparative research

A catalogue of challenges facing the use of diagnostic classification models to analyse large-scale assessment data.

Adam Coates

Introduction

Diagnostic classification modelling (DCM) is a statistical methodology used to estimate students' mastery of different skills (Rupp et al., 2010). In typical assessment practice, a test produces only a single, overall score for each student. This has limited value in providing direction for teaching and learning. In contrast, DCM can be designed to evaluate multiple skills, and therefore provide valuable information to direct future practice (Sessoms & Henson, 2018).

DCM are typically tied to cognitive diagnostic assessment. In the context of education, cognitive diagnostic assessment is an approach which aims to create tests which can measure specific skills so that feedback can be given about students strengths and weaknesses (Leighton & Gierl, 2007). This requires that tests are designed using a cognitive model of the knowledge and processes required to successfully complete each task (Leighton & Gierl, 2007; Zhang et al., 2024). DCM can then be used to analyze student responses to the test and estimate the skills they have mastered.

Although the primary function of DCM is to provide feedback for individuals, DCM have been used to analyze the learning across different countries by analyzing data from international large-scale assessments (ILSAs). The results produced by these studies demonstrate a potential value of analyzing ILSA data using the DCM methodology. However, ILSAs are not cognitive diagnostic assessments: they are not created based on a cognitive model. As a result, important issues can arise when analyzing ILSAs with DCM.

This study draws on previous research and an analysis of TIMSS 2023 grade 4 mathematics data to illustrate the challenges that can arise when using DCM to analyze large-scale assessment data. It identifies weaknesses in previous research that can be overcome (e.g. accurate estimation of error) and more fundamental issues that may not be able to be resolved.

Research objects and hypotheses

The fundamental objective of the study is to adapt the TIMSS methodology, which operates from an IRT framework, to be suitable for a diagnostic classification framework. In the process of doing this, key challenges are identified and possible routes to overcoming these are attempted. The presentation presents a catalogue of these challenges. In essence, the study hypothesizes that it is possible to apply the DCM framework with ILSA data, and the success of this application is tested along a variety of dimensions.

Data

Mathematics test response data from TIMSS grade four 2023 is used. Data from booklets 1-14 are analyzed, and all countries (57) are included, giving a total of 357,571 students.

To compare the issues illustrated by the present analysis with previous literature, a systematic search was conducted. Literature was searched on dimensions.ai, (Dimensions AI, n.d.); the *Educational Resources Information Center* (ERIC); the *British Education Index*; and *Education Source*. For the three education databases, the search terms were: *("cognitive diagnostic" OR "diagnostic classification") AND (math OR mathematics)*. Because *dimensions.ai* is not focused on educational research the additional search terms *AND (education OR student OR pupil OR school)* were used. Abstracts were screened to identify studies which used empirical data, then papers were read to identify those using data from international large-scale assessments. Overall, 54 studies were collected, of which ten used PISA data and the remaining 44 analyzed TIMSS results.

Method

The first step in diagnostic classification modelling is creating or selecting a cognitive model. This model represents a theoretical assumption about the cognition required to solve each item correctly. The model includes the identification of skills (i.e. knowledge and processes) and assumptions about how these skills combine when solving problems. The skills for the cognitive model were taken form the TIMSS framework for mathematics.

The basic TIMSS framework for grade four mathematics includes three cognitive skills (knowing, applying, and reasoning) and three content domains (number; geometry and measurement; and data). However, analysis shows that these skills are highly correlated (>.90), and this implies that the subscores will have little value. However, the TIMSS framework includes an expansion of the content domains: 1. whole numbers; 2. expressions, simple equations, and relationships; 3. fractions and decimals; 4. measurement; 5. geometry; 6. reading, interpreting, and representing; and 7. using data to solve problems. Each item in the assessment is classified as testing one content skill and one cognitive skill. A Q-matrix was produced based on the TIMSS allocation of skills to items.

The skills can be modelled as having independent contributions to the probability that a question will be correctly answered, an interaction, a compensatory relationship, or a combination of these relationships. In the absence of a reasonable theoretical assumption, models of different relationships were tried and the model showing best fit with the data was selected. This was the GINA model, which includes independent contributions and an interaction for the two skills in each question.

The DCM model was estimated and used to create plausible values following the TIMSS methodology to the extend possible. Learning progressions were produced based on two different approaches drawn from previous studies. The first approach involves identify the largest group of students who have mastered one skill, the largest group of those who have mastered two skills, etc. (e.g. as in Zhu, 2023). The second approach involves selecting learning states in order of size until a threshold is reached (e.g. 85-90% as in Dogan & Tatsuoka, 2008; Wu et al., 2020).

Results and findings

Key issues with using DCM with large-scale assessments are now presented in the sequence in which they arise when following the DCM methodology.

Cognitive model: The first issue with using DCM to analyze large-scale assessments is whether a meaningful cognitive model can be produced. Because large-scale assessments are holistic test of a subject, they span a wide range of topics. This limits how narrowly defined skills can be. Typically, skills in the cognitive diagnostic assessment framework are fine-grained, such as "converting an improper fraction to a mixed number." Skills at this level can be interpreted relatively easily as mastered (or not) since they refer to specific steps in the processes of solving in a problem. In contrast, because the content of large-scale tests is so broad, the skills in a cognitive model are equally broad. For example, many studies use TIMSS frameworks for the skills, meaning that *geometry* is a skill. Other research has created their own sets of skills by inspecting items. While the skills in these models are typically more specific than those in the TIMSS framework, they are typically composite skills, combining multiple individual sub-skills. For example, one skill in the framework of Lee et al. (2011) *calculating and estimating perimeters, areas, and volumes*. Because these broad and composite skills do not lie within the cognitive diagnostic assessment framework, theoretical work needs to be done to explain how mastery of these skills should be interpreted.

For this study, the TIMSS framework was used so that the complete sample of the 2023 cycle could be analyzed. This unavoidably carries issues with the interpretation of skills. A basic theoretical framework to explain the meaning of these skills can be produced, but space does not permit detailing it here.

Distinctiveness of skills: Another issue that can arise from analyzing holistic tests is that the skills may not be distinct. Because broad skills are necessary to cover the range of mathematical processes required to solve items, there can easily be strong correlations between the mastery of these skills. For example, the IRT produced subscores in the TIMSS framework have correlations over r=.95 (Camilli & Dossey, 2019). This means the individual subscores are of little value. Existing DCM studies have not reported the distinctiveness of the classifications they produce.

In this study, it was found that the basic TIMSS framework (cognitive domains and *number*, *geometry,* and *data*) produced high correlations between the skills (>.9). However, the slightly more fine-grained framework contained much lower correlations, with an average of .58 (e.g., with *fractions and decimals* as a skill).

Alignment between the cognitive model and the test: This is an issue which primarily affects studies using the theoretical framework of the test organization, but it can carry across to studies which develop their own cognitive models. In essence, there can be doubts that a cognitive model sufficiently and accurately describes the skills required for each item. For example, the TIMSS framework allocates one cognitive and one content skill to each test item. However, studies which use Q-matrix refinement approaches to estimate the allocation of skills to questions have found that multiple content skills might be used for some TIMSS questions (e.g. Delafontaine et al., 2022; Ma, 2019). This issue could be resolved my inspecting each question

and re-classifying them, allowing for multiple content skills; however, the issue that the skills used in a model are likely to be very broad can make this process challenging.
Because this study used the TIMSS framework, it inherits the potential misallocation of skills to items. Although empirically Q-matrix estimation could help identify items where these issues arise, without being able to inspect the items, it is arguably most reasonable to work with the TIMSS specification rather than blindly accept the statistical recommendation.
Model selection: From a more technical perspective, DCM studies using ILSA data face a problem that, in retrofitting a cognitive model, they are rarely confirmative. One of the nine characteristics in the definition of DCM by Gierl and Cui (2008) is that DCM are confirmatory in nature. However, the relationships between skills (independent, interactive, etc.) is not assumed in ILSA frameworks, nor do studies who create their own frameworks make these assumptions. Instead, studies have exclusively empirically-selected the relationship between skills for the cognitive model by analyzing data with various models and selecting the one with best fit. Because the cognitive models developed for ILSAs have broad skills, they cannot draw on existing cognitive science to guide the choice of relationships between skills. As mentioned earlier, this results from the issue that traditional cognitive models use particularly fine-grained skills.
Sample size: ILSAs engage in robust sampling methodologies. However, only a limited number of items in the assessments are released publicly. Theoretically, a study can request access to all items, analyze them, create a cognitive model, then analyze all student responses. In practice, studies have analyzed only one or two booklets, meaning that les than 20% of the complete ILSA sample is used. Moreover, although studies using ILSA frameworks as the basis for a cognitive model could analyze the complete ILSA sample, this is not done (e.g. Dogan & Tatsuoka, 2008; Liao et al., 2024; Wu et al., 2020). Additionally, studies rarely investigate learning across all countries in the sample: often only one or two countries are studied (e.g. Im & Park, 2010; Şen & Arican, 2015). By limiting the sample in a study to a small sub-sample of the data for an ILSA cycle, the accuracy and generalizability of the DCM results is negatively affected.
This study used the complete student response data for TIMSS 2023 mathematics fourth grade, and, therefore, addressed this issue.
Standard error estimation: ILSAs use plausible values to accurately estimate standard errors. This involves using information from background questionnaires in the estimation of students' ability. Previous studies have not engaged with this process, and only one study used Jackknife Repeated Replication. This means previous studies underestimate the level of variance in the population.
This study created a form of plausible values to address this shortcoming in previous research. However, the process of doing this is particularly challenging and may explain why it has not been used in previous research.

Conclusion

This study has identified a number of key challenges that can arise when using DCM to analyze ILSA data. It demonstrates that some of these challenges can be overcome, and these mostly relate to practical issues. However, important conceptual problems remain, and these can easily be interpreted to mean that the use of DCM with ILSA data is not wise. It is therefore recommended that the use of DCM is limited to contexts where a test is designed based on a cognitive model and suitably fine-grained skills can be used.

Keywords: ILSA, DCM, methodology

Trends in reading teacher shortage: Using 20 years of evidence from PIRLS

Nurullah Eryilmaz

Introduction

Teacher shortages are a pressing global concern with direct implications for achieving the United Nations' Sustainable Development Goal 4 (SDG 4): equitable access to quality education (UNESCO, 2024). While research on teacher shortages has often focused on STEM subjects or general workforce gaps (Fuentes & Bloom, 2023; Reinsfield & Lee, 2022), far less attention has been paid to reading teacher shortages—a critical issue in early education. This study investigates longitudinal trends in principal-reported concerns about reading teacher shortages across 20 years of PIRLS data (2001–2021). We focus particularly on disparities across socioeconomic status (SES) and geographic location (urban vs. rural), offering a comprehensive cross-national analysis of how these shortages are distributed and how they have evolved over time.

Research Object and Hypothesis

The study aims to determine:

1. How concerns over reading teacher shortages have changed over time;
2. Whether schools serving low-SES populations experience more acute shortages than high-SES schools; and
3. Whether rural schools report higher shortages than urban schools.
   Our hypothesis is that teacher shortage concerns are more prevalent in schools serving disadvantaged populations and that these disparities persist or widen over time.

Data

We use principal questionnaire data from the Progress in International Reading Literacy Study (PIRLS) across five assessment cycles: 2001, 2006, 2011, 2016, and 2021. The sample includes 66 countries, with 18 participating in all five cycles. The teacher shortage variable is derived from principal reports on the extent to which a shortage or inadequacy of teachers specialized in reading impacts instructional capacity. To facilitate analysis across cycles and countries, we harmonize item wording variations over time and dichotomize responses into "no shortage" and "shortage" categories. SES is operationalized based on the percentage of economically disadvantaged students reported by principals (Johansson & Myrberg, 2019), while school location is categorized using population thresholds (rural < 100,000; urban ≥ 100,000). The final analytical dataset includes over 500 country-year-subgroup observations.

Method

We apply fixed-effects regression models to examine within-country changes over time and isolate the role of year, SES, and location. The models estimate teacher shortage percentages for each country-year subgroup, weighted using PIRLS school-level sampling weights. We also conduct variance decomposition analyses to quantify the share of variance attributable to countries, years, and subgroup interactions.

Models are estimated separately for (1) national trends, (2) SES disparities, and (3) rural-urban gaps. Each includes interaction terms between year and subgroup variables to test whether disparities have shifted over time. Robustness checks are conducted on subsets of countries with consistent participation.

Results and Findings

RQ1 - Overall Trends:

Concerns over reading teacher shortages have increased modestly over time. In 2001, 52% of principals across countries reported no shortage, dropping significantly in subsequent cycles. The largest increases in concerns occurred in 2011 and 2016. However, variance decomposition reveals that 76% of the total variance is between countries, suggesting that national factors dominate.

RQ2 – SES Disparities:

On average, low-SES schools report higher concerns over reading teacher shortages than high-SES schools, though these gaps are not always statistically significant across years. Countries like the United States and New Zealand show growing gaps (García & Weiss, 2020), while others like Sweden and Hungary show more equitable or reversed trends over time. Variance decomposition supports context-specific patterns.

RQ3 – Urban-Rural Disparities:

Rural and urban differences in teacher shortage concerns are generally smaller and less stable. In some countries (e.g., the U.S.), rural schools reported more shortages in specific years, while in others (e.g., New Zealand), urban schools became more affected over time (Ingersoll & Tran, 2023). The fixed-effects model shows no significant overall location effect.

Additional Findings:
Bulgaria stands out as a success story, reducing shortage concerns by over 40 percentage points from 2001 to 2021—potentially linked to policy reforms following EU accession (Aykac & Sahin, 2018). Conversely, the U.S. and the Netherlands experienced rising concerns. Some countries (e.g., Scotland, Latvia) maintained stable, low shortage levels across cycles.

Findings of Italy
Italy's trends in principal-reported reading teacher shortages show moderate and relatively stable levels across PIRLS cycles from 2001 to 2021. There was a slight increase in 2011, followed by a return to previous levels in 2016 and 2021. While Italy does not exhibit extreme fluctuations compared to countries like the United States or Hong Kong, it maintains a steady mid-range level of concern. Importantly, this overall stability masks significant variation across school contexts, particularly with regard to socioeconomic status.

Across all cycles, low-SES schools in Italy consistently report higher levels of teacher shortages than high-SES schools, with the gap most pronounced in 2011 and again in 2021. This suggests persistent inequalities in access to specialized reading teachers, reinforcing concerns about equity in the Italian education system. Geographic disparities—between rural and urban schools—are less consistent. While rural schools reported more shortages in 2011 and urban schools did so in 2016, by 2021, the gap between the two had largely closed. These findings indicate that while overall shortages remain stable, Italy faces challenges in ensuring equal distribution of qualified reading teachers across socioeconomic groups.

Discussion
The findings suggest that teacher shortages, particularly in reading, are not universally experienced or inevitable. While some countries have succeeded in reducing shortages, others have seen increases, highlighting the role of national policy contexts. The variation between countries suggests that international assessments like PIRLS can serve not only for benchmarking but for identifying effective practices and understanding system-level dynamics (Boeren, 2019).

Disparities in shortage concerns by SES underscore the relevance of Bourdieu's theory of social reproduction (Bourdieu, 1973; Nash, 1990): students in low-SES schools often face more limited access to qualified teachers, perpetuating inequality. However, the presence of countries that have narrowed these gaps points to policy levers that can interrupt this cycle.

Implications
This study provides three key insights:

1. National Context Matters – Cross-country differences dominate, suggesting country-specific interventions are necessary.
2. Equity Is Variable – While average trends show weak effects, country-level analysis reveals substantial variation in how teacher shortages affect disadvantaged schools.
3. Longitudinal Assessment is Critical – One-year snapshots are insufficient; sustained monitoring over time is necessary to identify systemic problems and successes.

Countries such as Bulgaria, Sweden, and the Slovak Republic offer models for reducing shortages or redistributing teaching capacity more equitably. Future research could investigate the timing and nature of policy shifts, such as Bulgaria's teacher education reforms or Sweden's incentive programs for difficult schools.

Conclusion
Teacher shortages—especially of specialized reading teachers—remain a global concern with substantial national variation. This study, based on 20 years of PIRLS data, shows that concerns have increased in some countries, especially in low-SES schools, while others have made measurable progress. Addressing this issue requires targeted, evidence-based policies and a commitment to long-term monitoring. International assessments like PIRLS can play a pivotal role in helping systems move from comparison to transformation—supporting education systems to not only measure outcomes but to understand and improve the conditions that shape them.

Keywords: Teacher shortages, PIRLS, reading teacher specialization, educational inequality, international comparison

Interaction between ICT Use and Academic Performance and Educational Equity in the Italian Context

Marina García Herrero – Fernando Martínez Abad – María José Rodríguez Conde

Introduction

Fair access to quality education becomes a crucial issue for long-term development in the current environment of swift technological advancement and growing global interconnectedness. In this context, the United Nations' 2030 Agenda's Sustainable Development Goal 4 (SDG 4) highlights the importance of facilitating equitable participation in affordable vocational education, mitigating gender and socioeconomic inequalities, and ensuring universal availability of high-quality higher education (United Nations, 2015).

In this regard, Large-Scale Assessments have become key instruments for measuring students' academic performance and analysing the factors associated with this performance in both national and international contexts. These assessments enable researchers to make comparisons between countries with the aim of finding out what is being done well in certain contexts and trying to replicate these good practices in others to improve education (Frade-Martínez et al., 2024). In addition, large-scale assessments have also become essential for obtaining in deep knowledge about different issues as educational equity, highlighted by the Organisation for Economic Co-operation and Development as a fundamental value and aim of education policy, as well as an ethical principle linked to the idea of justice, which states that everyone should have the chance to reach their full potential, regardless of their background (OECD, 2023). This statement corresponds to the proposal of equity as equality of opportunity from Enchikova et al. (2024), who point out that the term "equality of opportunity" is referred to "the connection between background characteristics and educational results" (p.4).

Within this framework, the role of digital technology in education has gained increasing attention as both a potential driver and barrier to equity. According to the Global Education Monitoring Report 2023 (2023), although digital technology has transformed education and offers substantial potential, its positive impact relies on equitable access, clearly defined goals and principles, effective support, and appropriate adaptation to diverse contexts. As such, opinions vary widely on the extent to which digital technology can truly enhance the quality of education.

Recent international assessments offer empirical evidence to explore this intersection between ICT use and academic achievement. Thus, data on average OECD countries from PISA 2022 assessment reveal that the relationship between students' use of digital devices at school and their mathematics performance and well-being varies depending on the purpose of use (OECD, 2024a). Specifically, moderate use of digital devices for learning activities is associated with higher academic performance and a stronger sense of belonging at school, whereas students who spent more than one hour using digital devices for leisure purposes tended to achieve lower scores in mathematics and reported a diminished sense of belonging at school. These findings highlight the need to further explore the implications of ICT for the student's learning achievement and opportunities in different educational contexts.

Research Object And Research Questions

Based on the issues outlined in the previous section, the study focuses on the Italian context and addresses the following research questions:

RQ1. What type of association exists between the use of information and communication technologies (ICT) and academic performance in Italy, compared to the benchmark[6]?

RQ2. What is the effect of student SES on the relationship between achievement and ICT factors (equality of opportunity) in Italy, in comparison with the benchmark?

Considering these questions, the main objective of the study is to examine the extent to which the use of information and communication technologies (ICT) in the Italian educational context influences students' mathematics performance and educational equity, by comparing these results with the average of countries included in PISA study.

Data

We used PISA 2022 data, specifically a list of countries which included all the variables used. Thus, countries were excluded for these reasons:

1. Didn't include any of the ICT scales used as explanatory variables.
2. Didn't include the school ownership information.

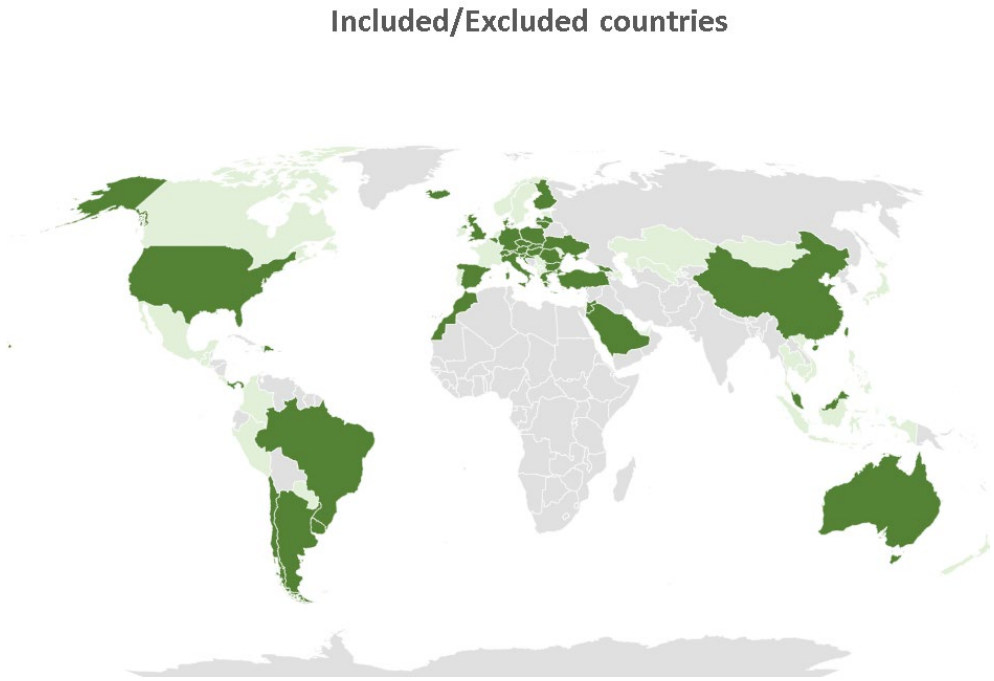
[6] We use the term "benchmark" to refer to the international standards in the obtained indicators (OECD, 2023).

Method

We drove an ex-post facto secondary data analysis, with an explanatory and correlational approach. Figure 1 shows the list of PISA 2022 countries included (dark green) and excluded (light green) in the analysis.

Figure 1

Countries included in the benchmark model



The included variables were:

- Criterion variable: Math achievement
- Explanatory variables
    - Contextual factors: gender (0=Male; 1=Female), socioeconomic status (ESCS), immigration status (IMMIG: 0=1st generation immigrant; 1=2nd generation immigrant or native); Years of early childhood education (DURECEC), language at home (0=Different from assessment; 1=The same as assessment), size of school location (SCHLOCATION), school ownership (SCHTYPE: 1=Public; 0=Private).
    - ICT factors (frequency of use): ICT use in subjects inside school (ICTSUBJ), ICT use in enquiry educational activities inside school (ICTENQ), students' feedback received with ICT (ICTFEED), ICT use outside school for educational activities (ICTOUT), ICT use for leisure purposes in weekdays (ICTWKDY), ICT for leisure purposes during the weekend (ICTWKEND).

Regression multilevel models (L1=Students; L2=Schools) were computed in the included countries, all together (benchmarking model) and separately by country. Multilevel models were estimated with random intercepts and fixed slopes and included the interactions between SES and ICT factors (equality of opportunity estimation). As recommended in PISA 2022 technical report (OECD, 2024b), the models included: combined student and school weights; replicate weights; stratum information; and plausible values (PV). We computed 10 independent models, one for each PV. Then, to estimate the final models' parameters and standard errors (SE) we computed, respectively, the parameter averages and the SE with Rubin's formula (Rubin, 1996).

We used R software to compute the multilevel models and to obtain the related graphs. First, we used package *mice* to replace missing values separately in each country with multiple imputation. After that, we used the package *survey* to define the PISA 2022 sample design, and the package *svy2lme* to compute multilevel models taking into account the sample design (including weights, replicates and strata). Finally, we used the packages *ggplot2* and *interactions*, respectively, to obtain the comparative graphs of the

parameters obtained by country and the graphical representation of the significant interactions between variables in the Italian sample.

Results And Findings

Multilevel Regression Model

As presented in Table 1, three contextual variables in Italian model showed statistically significant p-values: ESCS, with a large positive coefficient; GENDER, with a moderate negative beta; and LANG, with a moderate positive parameter. Notably, the positive association between ESCS and mathematics academic performance was weaker in Italy compared to the benchmark, suggesting that socioeconomic status plays a less influential role in the Italian context. Conversely, the negative effect of GENDER was more pronounced in Italy, indicating a stronger gender gap, with women underperforming compared to men. In contrast, the effect of LANG was considerably stronger in Italy, implying that the language spoken at home has a greater impact on academic performance compared to the average benchmark.

In terms of ICT factors, four ICT-related variables were significantly associated with academic performance. ICTSUBJ had a positive beta coefficient higher than that of the benchmark, suggesting that a higher frequency of ICT use in school subjects is strongly linked to academic achievement in Italy. ICTFEED and ICTWKDY showed moderate negative effects, very similar with the benchmark countries, suggesting that receiving feedback via ICT and an excessive leisure-related ICT use may be associated with lower academic achievement. In contrast, ICTOUT showed a positive effect of the ICT use outside school for educational activities, though weaker than in the benchmark group, indicating a relatively lower influence in the Italian context.

Regarding the interactions, two statistically significant parameters were identified in Italian database. On the one hand, the interaction ESCS*ICTSUBJ revealed that the positive relationship between ICT use in subject learning and academic performance is stronger among students with higher SES. On the other hand, the interaction ESCS*ICTFEED indicated that the negative association between ICT-based feedback and academic performance is stronger for students with higher SES.

Table 1

Multilevel Regression Model

	b parameter		**SE**	**t**	**p**
	Benchmark	**Italy**			
Intercept	430.779	442.415	38.056	11.625	<.001
ESCS	37.687	26.392	2.122	12.440	<.001
GENDER	-11.507	-18.370	3.688	-4.982	<.001
IMMIG	-1.829	4.045	9.455	0.428	.669
LANG	4.333	16.142	5.504	2.933	.003
DURECEC	6.774	0.262	1.685	0.155	.877
SCHLOCATION (L2)	8.699	3.319	5.489	0.605	.545
SCHLTYPE (L2)	-10.361	3.762	28.259	0.133	.894
ICTSUBJ	5.047	8.740	2.052	4.259	<.001
ICTENQ	-1.858	2.774	2.101	1.320	.187
ICTFEED	-14.038	-14.168	1.931	-7.337	<.001
ICTOUT	15.889	10.906	1.934	5.639	<.001
ICTWKDY	-15.336	-15.128	3.079	-4.914	<.001
ICTWKEND	-3.900	-3.069	2.611	-1.175	.240
ESCS*ICTSUBJ	2.239	3.689	1.770	2.084	.037

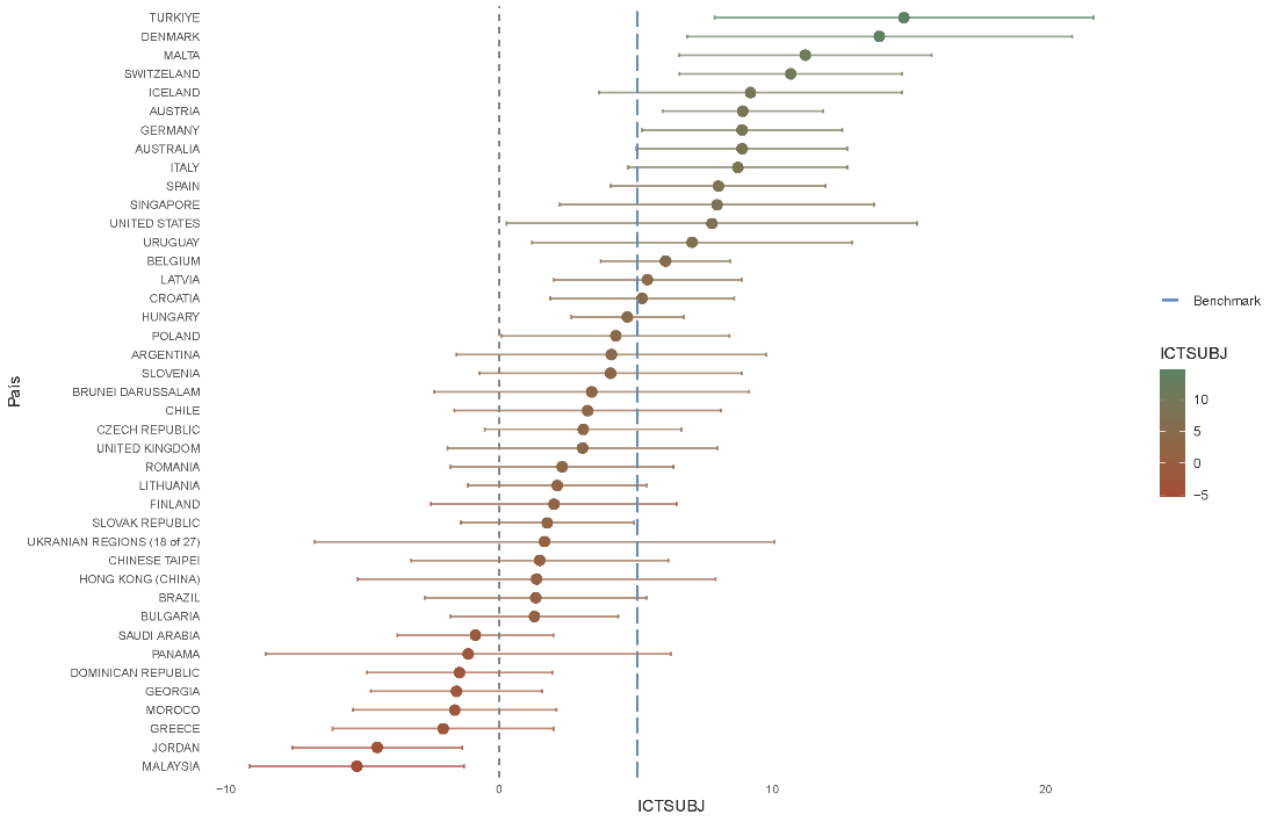
ESCS*ICTENQ	1.196	1.938	2.239	0.865	.387
ESCS*ICTFEED	-3.004	-5.760	2.102	-2.740	.006
ESCS*ICTOUT	1.634	-1.629	2.499	-0.652	.514
ESCS*ICTWKDY	-5.295	-2.259	2.898	-0.779	.436
ESCS*ICTWKEND	-1.293	-4.726	2.883	-1.639	.101

Interval charts by country

As for the ICTSUBJ variable (Figure 2), the statistically significant positive effect shown by Italy is slightly higher compared to the benchmark. However, the effect achieved in Italy is not significantly higher than the benchmark. Turkey, Denmark, Malta, Switzerland, Austria and Germany show stronger positive effects, above the benchmark. In contrast, Malaysa and Jordan are located at the lower end, suggesting that the use of ICT in the subjects is associated with lower academic achievement.

Figure 2

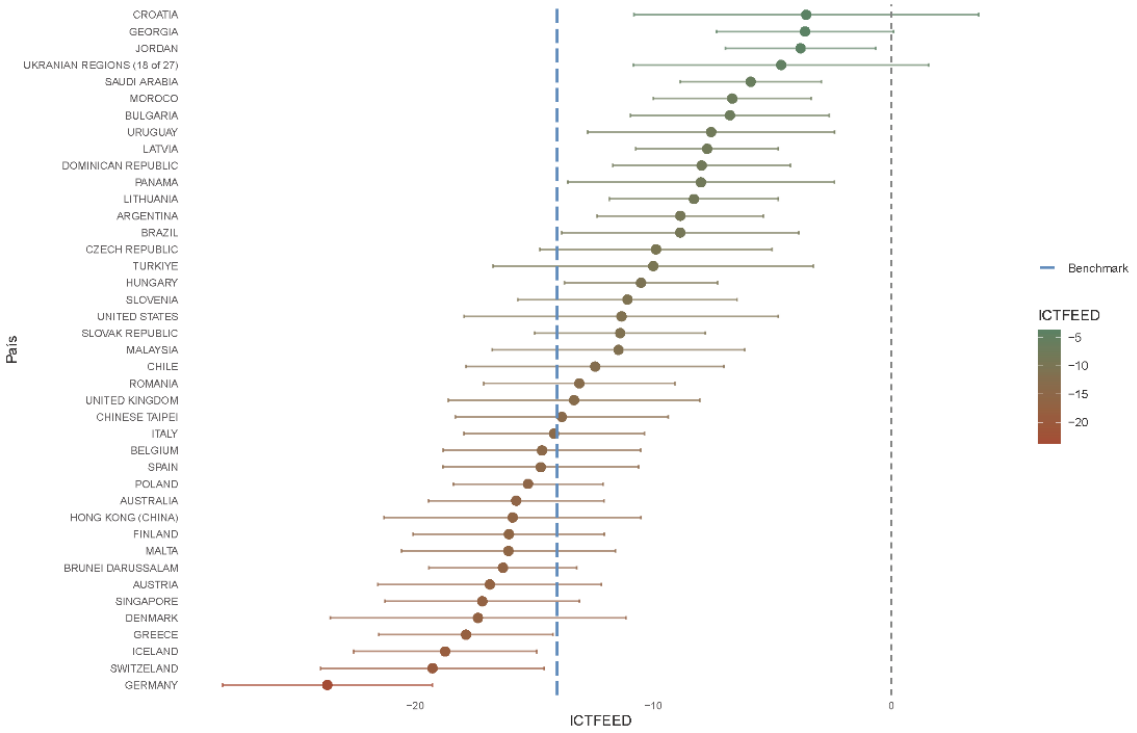
Interval charts by country (ICTSUBJ



Regarding ICTFEED (Figure 3), the statistically significant negative effect shown by Italy is very close with the benchmark. Other countries with negative significant effects below the benchmark are Germany, Switzerland, Iceland and Greece. On the opposite, among the countries with positive effects exceeding the benchmark we find other European countries like Croatia, Georgia, Ukraine, Bulgaria, Latvia and Lithuania, indicating that in these contexts the use of ICT for receiving feedback is associated with better mathematics' academic achievement.

Figure 3

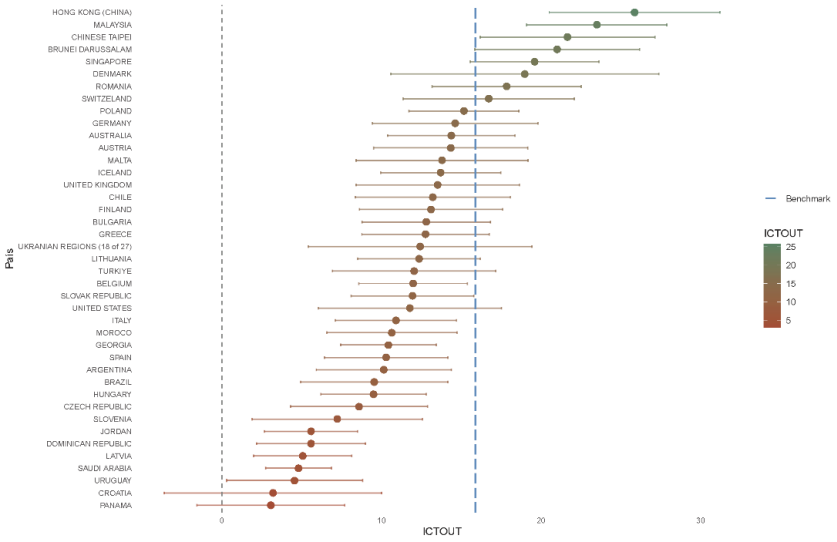
Interval charts by country (ICTFEED)



The ICTOUT variable (Figure 4) achieved in Italy a statistically significant positive effect, in line with the benchmark. Other countries with positive effects, but above the benchmark, are Asian contexts such as Hong Kong (China), Malaysia and Taiwan. On the other hand, countries without significant effects in the ICTOUT variable, below the benchmark, are Croatia and Panama.

Figure 4

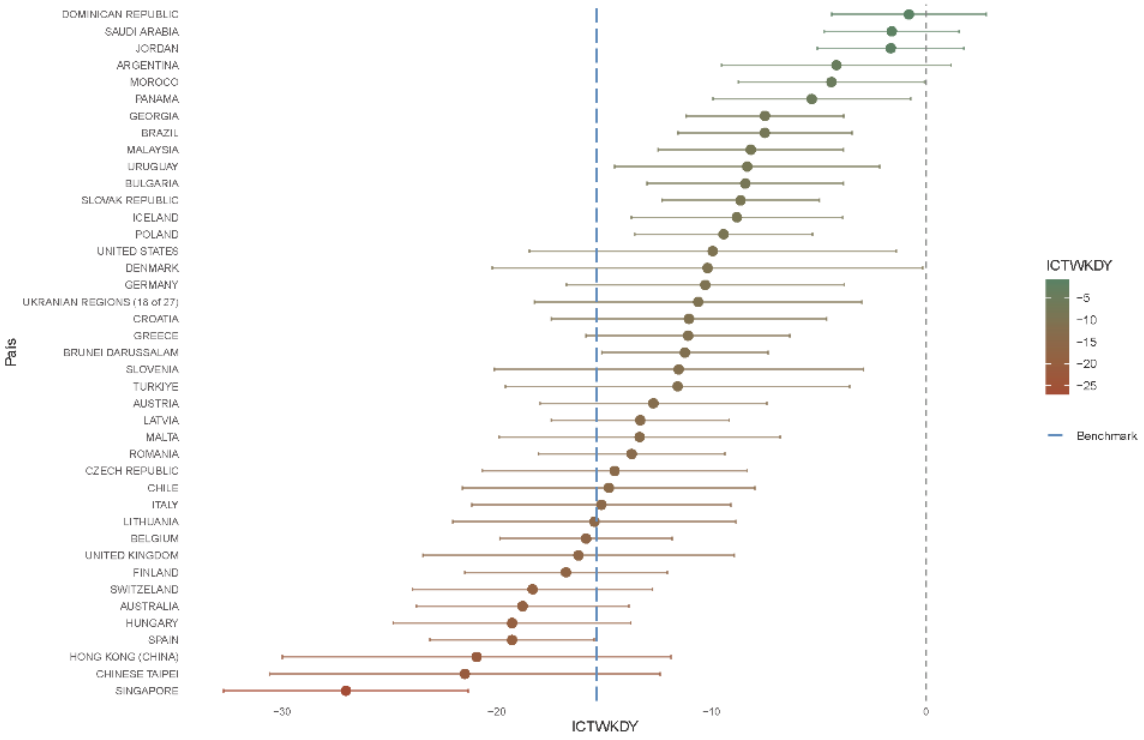
Interval charts by country (ICTOUT)



Finally, in the variable ICTWKDY (Figure 5), Italy showed a statistically significant negative effect, at the benchmark level. With non-significant scores above the benchmark, there are some low- and middle-income countries: Dominican Republic, Saudi Arabia, Jordan, Argentina and Morocco. At the lower end, Spain and Singapore stand out for their significant negative scores above the benchmark.

Figure 5

Interval charts by country (ICTWKDY)

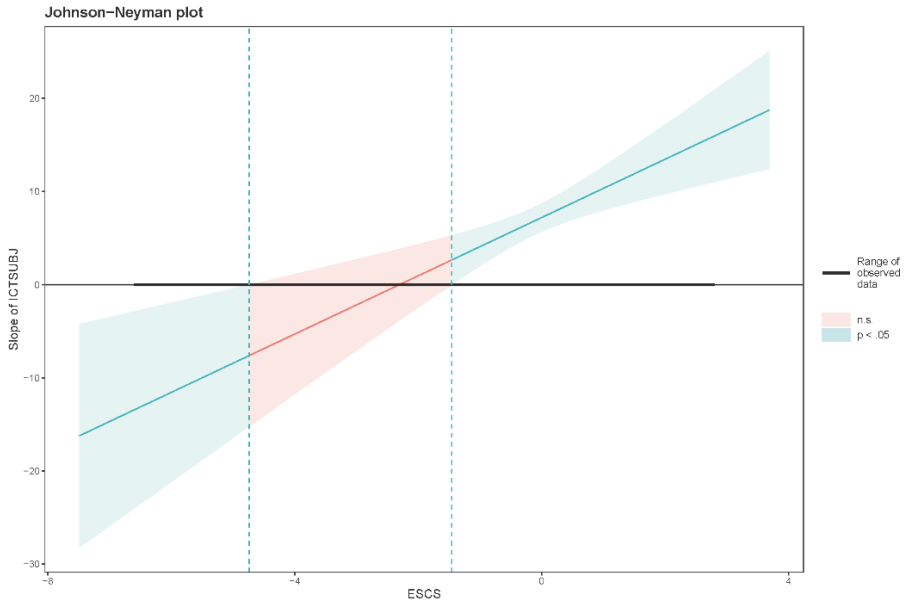


Significant interactions between variables

With respect to the interaction between ESCS*ICTSUBJ, as presented in Figure 6, it is observed that using ICT more frequently as an educational tool within subject activities is linked to better academic performance for students with medium to high ESCS levels, with the strongest effect at high levels. However, for students with very low ESCS, this relationship changes and ICT is inverse. This suggests that students from more advantaged socioeconomic backgrounds use ICT more effectively to support learning, whereas those from more vulnerable socioeconomic contexts may find the use of such tools distracting, moving away from their main objective, which is the improvement of their learning.

Figure 6

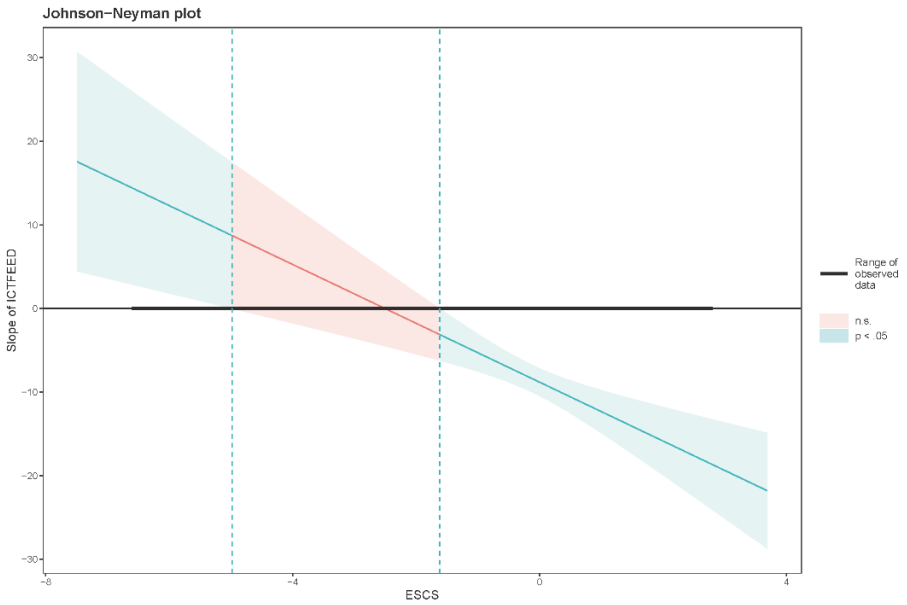
*Significant interactions between variables in the Italian sample (ESCS*ICTSUBJ)*



Regarding the interaction ESCS*ICTFEED, it is observed in Figure 7 that for students with medium-low to high ESCS levels, receiving more frequent feedback through ICT in educational activities is associated with lower academic performance, with this association being stronger for students with high ESCS. On the other hand, for students with very low ESCS, receiving feedback via ICT appears to be slightly associated with better performance, although this trend is weak since the confidence interval does not move far from zero. Therefore, this finding suggests that ICT-based educational feedback may act as a protective factor for equity, and it is recommended to promote this type of feedback especially among students from socioeconomically vulnerable backgrounds.

Figure 7

*Significant interactions between variables in the Italian sample (ESCS*ICTFEED)*



Acknowledgements

- Grant PID2021-125775NB-I00 funded by MCIN/AEI/10.13039/501100011033/ and by ERDF A Way of Making Europe.
- Grant FPU23/02626 funded by Ministry of Science, Innovation and Universities (Spanish Government).

Keywords: PISA, ICT, equity, Italy

The Effect of Early Home Learning Activities on Achievement at the End of Primary Education

Alec Kennedy – Rolf Strietholt – Nina Hogrebe – Andres Strello

Theoretical framework
Early parental involvement in home learning activities (HLA) is widely regarded as a key driver of children's cognitive development. The literature suggests that HLA—including shared reading, storytelling, and number-related play—fosters early literacy and numeracy skills. However, isolating the causal effects of HLA is challenging due to endogeneity: parental practices are influenced by socioeconomic status, parental education, and children's traits.

Research objectives/questions
This study investigates: (1) whether early home learning activities predict fourth-grade reading and math achievement, and (2) how robust this relationship is across diverse socioeconomic contexts and countries. We pay particular attention to both literacy- and numeracy-related HLA, and examine whether unequal access to such practices contributes to achievement gaps.

Method of using INVALSI data
While INVALSI data are not directly used, the study relies on PIRLS (2001, 2006, 2011, 2016, 2021) and TIMSS (2011, 2015, 2019) datasets that share many structural similarities and design principles with INVALSI. Our method exploits within-student variation in early HLA to control for unobserved heterogeneity, using fixed-effects models on international large-scale assessment data from 35 countries.

Results
We find consistent positive associations between early HLA and later academic performance in both reading and mathematics. These relationships are robust across model specifications and persist when controlling for parental education and socioeconomic status. The effects are particularly pronounced in disadvantaged groups where early HLA is less prevalent, suggesting a policy lever for equity.

Relevance to the Seminar
Our research aligns with the seminar's focus on leveraging large-scale data for educational research and policy. The findings underline the importance of supporting early learning environments and show how international assessments can provide insights relevant to national policy agendas like those shaped by INVALSI.

Keywords: Early home learning, causal effect, social inequality

SESSION 7. INEQUALITIES IN EDUCATIONAL ATTAINMENT AND THE ROLE OF TEACHERS AND SCHOOL PRINCIPALS

ORGANIZER: UNIVERSITY OF PIEMONTE ORIENTALE
COORDINATOR: GIORGIA CASALONE - ALESSANDRA MICHELANGELI - JURGENA MYFTIU
21ST NOVEMBER: 2.00 P.M. - 4.00 P.M. {AULA MAGNA – RESEARCH 14}

Educational segregation in upper secondary schools: a school data-based analysis

Lorenzo Luatti – Sibilla Maria Filippi

Introduction

Lower secondary school (11-13 years old) represents the final stage of the Italian school system where social classes, gender, and citizenship backgrounds are still relatively mixed. After this stage, a more or less marked differentiation begins, giving rise to various forms of ***educational segregation***. As a result, there are upper secondary schools whose student composition - predominantly or entirely - is shaped by a combination of several factors: a) the score (high/low) obtained in the lower secondary final exam (esame di stato del primo ciclo), b) early grade overage (academic delay - one or more years behind the expected age-grade level), c) gender (male/female), d) socio-economic and cultural background, e) Italian or non-Italian citizenship (Fondazione Cariplo 2023; INDIRE 2023).

Educational segregation, in turn, is both connected to and reinforces: *inequalities in learning and educational outcomes* (INVALSI 2024; Fondazione Cariplo & Fondazione Rocca 2025; Schleicher 2020), 2) *early role assignment* based on widespread stereotypes in society and the labor market (Biemmi & Leonelli 2017), and thus contributes to shaping an upper secondary school system structured around tracks perceived (by teachers, families, and public opinion) as first-tier, second-tier, or even third-tier. Alternatively, these tracks may be seen as intended for capable or less capable students, natives or immigrants, boys or girls.

Both explicit and implicit forms of student selection are at play - upon entry and even during the course of study - often justified by "academic performance", though actually also linked to social, cultural, or implicit criteria.

The *educational segregation* that emerges after lower secondary school serves as a litmus test for the *residual early-selective orientation* of secondary education, which remains far from meeting the principles of equity, inclusion, and effectiveness set out in Articles 3 and 34 of the Italian Constitution (as well as Article 1, which affirms the equal dignity of all persons) (Benadusi & Giancola 2021; Biondi Dal Monte & Frega 2024).

This contribution aims to offer an overview of these phenomena and dynamics, drawing on a variety of sources and data analyzed through a diachronic-longitudinal, territorial, and upper secondary education lens. The paper also puts forward policy and action proposals.

The topic is addressed using a methodology for extracting, processing, and analyzing data, aimed at observing school disengagement and risk of dropout at the local level. This approach was developed and tested by Oxfam Italy across five school systems (Genoa, Turin, Savona, Vercelli, and the province of Arezzo: Luatti 2024; Luatti & Filippi 2025a) and, vertically, in fifteen Italian schools across both lower and upper secondary levels.

Research Focus, Objectives, and Hypotheses

This study focuses primarily on the lower and upper secondary school levels. Its goal is to provide, through an intersection of data on both explicit and implicit school disengagement (*dispersione scolastica* intended also as risk of school dropout), a detailed picture of inequality in learning and academic achievement, and how structural and contextual factors interact to shape educational outcomes.

Among the research hypotheses under investigation are the following:

- The "ambivalent" relationship between explicit and implicit school disengagement (*dispersione scolastica esplicita e implicita*): does reducing explicit school disengagement increase the implicit one (and vice versa)? Does minimizing predictive events of explicit school disengagement (failing grades,

repeated years, delays, school transfers...) - and thus the risk of early school leaving - necessarily imply lowering the bar in terms of learning and academic achievement?

- A broad and diverse range of data - highlighted in this contribution - confirms the "pyramidal" structure and the persistent selective orientation of the upper secondary school system (from the various types of *licei*, even internally, to technical and professional schools). The research hypothesis is that this structure contributes to significant consequences both ex ante and ex post: on school guidance decisions and the configuration of the labor market.
- Are educational segregation determined at entry and consolidated over time? Do the data confirm this hypothesis? What is the relationship between segregation and school dropout? What patterns can be observed at the territorial level? The hypothesis is that the "positioning" of each school within the system contributes to reinforcing both segregation and dropout phenomena.

Data Sources

The data sources used in this paper include:

A. The "Portale Unico dei Dati della Scuola" (PUDS): offers national datasets with historical series (since 2015/16), enabling diachronic-longitudinal analysis of state and parochial schools, from preschool to both educational cycles, disaggregated by grade level, school, school autonomy, and territory.
B. "Scuola in Chiaro" (SC): does not offer open-format data. For this paper, we include case studies from specific Italian cities and regions.
C. "INVALSI Statistical Service" (SSI): provides datasets allowing for diachronic-longitudinal analysis across various school levels, territories, and secondary education segregation. This paper draws on the 2024 data and historical series from the following reports: "General Report," "Provincial Levels Report," and "General Report with Dropout and Excellence Data," focusing primarily on grades 8, 10, and 13.

Specifically, we use:

- Student demographic data: number of students, classes, schools, students by gender and citizenship status (PUDS);
- Data on school path discontinuity and implicit school disengagement: school delay (PUDS), mid-year transfers (SC), basic skills proficiency levels and excellence/dropout patterns by grade (SSI);
- Data on school composition and tracking: schools with high concentrations of non-Italian students (PUDS), secondary school tracks by gender composition (PUDS);
- Data on transition phases: type of upper secondary school recommended in guidance reports (SC), student choices relative to lower secondary school final grades (SC), guidance report and admission to second year of upper secondary school (SC).

Methodology

This research adopts a quantitative approach grounded in rigorous methodological standards (Coe, Waring, Hedges & Ashley 2021), involving extraction, processing, and interpretation of open educational data. The methodology has been tested and made explicit in previous works (Luatti 2024; Luatti & Filippi 2025a; Luatti 2025b), proving effective for conducting analytical investigations into both explicit and implicit dropout in defined territories and school systems (e.g., districts, cities, provinces).

The approach is diachronic-longitudinal, applied to various geographic areas (Northwest, Northeast, Center-South, Islands) and different educational pathways (academic, technical, vocational tracks).

Our contribution will relate multiple analytical dimensions:

A. Trends in implicit school disengagement and excellence (SSI) by school type at grades 8 and 13 (natives and foreigners, male and female, by type of secondary school at grade 13);
B. Trends in proficiency in four basic competencies (SSI) at grades 8, 10, and 13 (natives and first-/second-generation immigrants, male and female, by type of secondary school at grades 10 and 13);
C. Learning progression or stagnation in the four basic competencies (SSI) from grade 8 to 13 over two five-year periods (e.g., INVALSI 2017–18 at grade 8 vs. 2022–23 at grade 13; INVALSI 2018–19 vs. 2023–24), including intermediate results at grade 10. Variables considered include citizenship (native and immigrant), gender, school type, and geographical area;

D. Interrelation between the above findings and data on explicit dropout factors (e.g. overage grade, transfers, final exam scores, guidance recommendations, concentrations of foreign students).

Results

The school data-based analysis allows for quantification and critical reflection on significant and persistent phenomena of educational tracking, emphasizing the need for progressive interventions to overcome them. It also reveals, through a multifaceted data approach, how tracking and both explicit and implicit school disengagement (*dispersione esplicita e implicita*) are interrelated, with varying configurations and outcomes depending on the variables considered.

The findings highlight the need for intervention on multiple levels: individual students (with tailored support, especially for those at risk), classrooms, schools or school networks, and the broader local and national education systems. Raising the quality of teaching and learning for all must be a priority. The heuristic potential of school data-based analysis and its possible benefits could be greatly enhanced by the development of software and applications to facilitate more user-friendly data extraction and processing, as well as by greater availability of open-format data on currently inaccessible variables from the two ministerial portals. Similarly, the analysis underscores the need for further disaggregation of INVALSI data (e.g., learning levels at grade 13 by type of upper secondary school).

Keywords: Inequalities in educational outcomes, **educational segregation,** territorial disparities, educational transitions

Lost Learning and Regional Resilience: The Impact of COVID-19 on Educational Inequality in Italy

Alessandra Michelangeli – Giorgia Casalone – Kateryna Tkach

Although education is recognized as a fundamental right in the Italian Constitution, the Italian education system continues to exhibit persistent territorial disparities in learning outcomes, rooted in structural socio-economic inequalities. The COVID-19 pandemic exacerbated these gaps, disproportionately affecting students from disadvantaged backgrounds and from specific regions. This study contributes to the theoretical debate on educational inequality in times of systemic shocks, with particular attention to the concept of territorial resilience and the adaptive capacities of schools and families.

The paper aims to: (i) Assess the impact of the COVID-19 pandemic on educational inequality in Italy across the first two educational cycles (primary and lower secondary); (ii) Measure the educational resilience of Italian regions during and after the pandemic; (iii) Identify the socio-demographic factors that most contributed to widening or narrowing educational gaps.

The analysis draws on individual-level data from INVALSI standardized tests, comparing results from pre- and post-pandemic years, with a particular focus on 2021. Educational inequality is measured using Generalized Entropy Indexes (GE), which are decomposed by gender, parental education, citizenship, and region. This approach allows the identification of both within-group and between-group components of inequality.

Findings show that stronger educational performance is concentrated in economically advantaged regions, while areas with higher socio-economic marginalization — including a greater presence of low-income households, single-parent families, and foreign-born residents — display significantly weaker outcomes. Some regions, however, exhibited signs of educational resilience, suggesting that local policies and contexts can mitigate the negative effects of crisis-induced disruptions.

This work offers a data-driven contribution to the seminar's themes by providing a quantitative assessment based on INVALSI data. It highlights the evolving patterns of educational inequality in Italy and the role of territorial context in shaping responses to systemic shocks. The findings inform the design of spatially targeted educational policies aimed at reducing inequality and fostering equity in learning outcomes.

Keywords: Education, inequality, resilience, Invalsi data**.**

Understanding the Gender Gap in Mathematics through the Didactic Contract: Insights from INVALSI Assessments

Serena Monica

Rationale and Theoretical Framework

Despite recent progress in female representation within STEM, significant gender disparities persist in both education and employment. These imbalances stem from historical exclusion and are influenced by internal factors, such as math anxiety and low self-efficacy (Mazza & Gambini, 2023), and external socio-cultural factors like stereotypes and educational policies (Giberti, 2019). Cultural context plays a crucial role in shaping gender differences in mathematics performance (Pajares, 2005). While complete gender equality remains a challenge, promoting equity in education is widely recognized as essential. International bodies like UNESCO emphasize gender equality as a core educational goal, and large-scale assessments continue to monitor andinvestigate these disparities across different educational systems (UNESCO, 2015 unesdoc.unesco.org/library/). In this context, the present study aims to highlight the presence of a gender gap in relation to specific mathematics items from the INVALSI assessments and to explore potential connections with key theoretical frameworks in mathematics education (Ferretti & Giberti, 2021).

In this research, the primary theoretical lens adopted is the didactic contract (Brousseau, 1986). The didactic contract refers to the implicit expectations between teachers and students within a shared instructional framework. Recognized as a robust construct in mathematics education (EMS-EC, 2011), it helps explain classroom dynamics and student behaviour. Research has shown that gender disparities in mathematics may be linked to the didactic contract (Ferretti & Giberti, 2021), particularly the tendency of female students to adhere more strictly to its norms. A significant component of this construct is the clause of formal proxy (D'Amore, 2008), which arises when the resolution of a problem is entirely entrusted to calculations, without a critical interpretation of the process and its results.

Research questions

1. How do specific mathematics items in large-scale assessments reflect patterns of gender-based performance differences?
2. What insights do students' reasoning patterns in the qualitative analysis highlight about the relationship between the gender gap and the didactic contract?

Methods

The study adopts a mixed-methods design (Johnson & Onwuegbuzie, 2004), structured as QUAN→QUAL. It begins with a quantitative analysis of standardized INVALSI test data from a nationally representative sample of fifth-grade students (INVALSI Statistical Service), followed by a qualitative interpretation of open-ended responses of 30 students of the last year of the primary school. The selected mathematics item, the 5th grade item 9 of the 2017 INVALSI test from the "numbers" domain (Figure 1), showed a high percentage of incorrect responses, approximately 60% (Figure 2).

Figure 5: Item 9, year 2017, source www.gestinv.it (translated by the author)

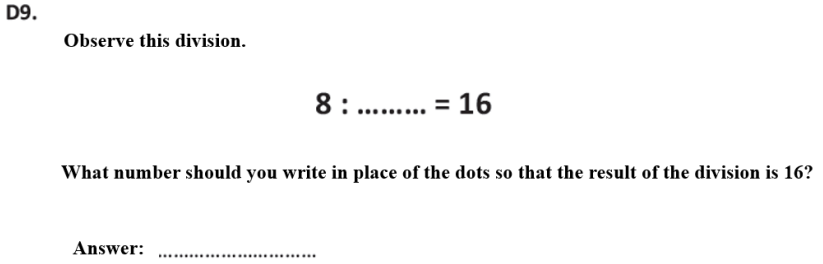
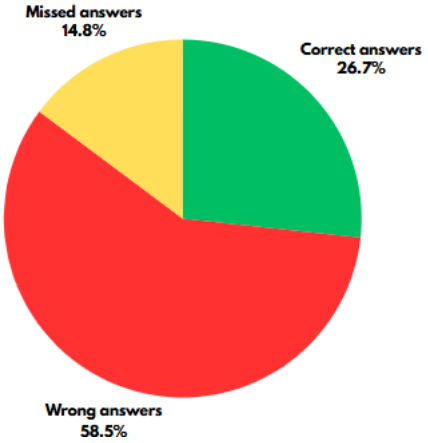
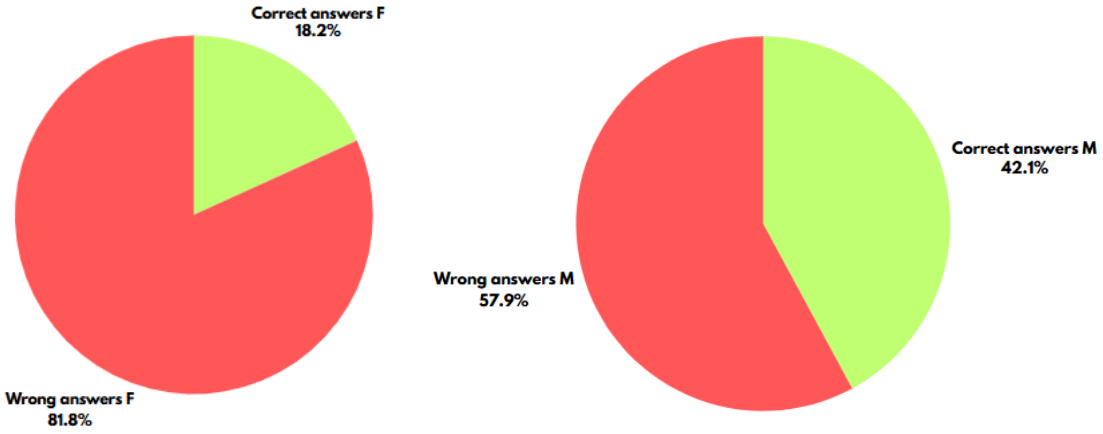


Figure 6: Percentages of item 9, year 2017, source: www.gestinv.it (translated by the author)



Furthermore, some gender differences have also been highlighted, confirmed through a two-proportion Z-test (Figure 3). The qualitative analysis, on the other hand, was conducted by the administration of a questionnaire in which was reported the item and a space for the explanation of the reasoning. In the study the open-ended answers were classified within the frameworks of mathematics education, with particular emphasis on the didactic contract.

Figure 7: Percentages of females (on the left) and males (on the right) correct answers, item 9, year 2017



Results:

A two-proportion Z-test was carried out, with the software R, to examine gender differences in students' responses to a mathematics item. The analysis compared the proportion of correct answers between female and male students. Results indicate a statistically significant disparity in performance, with female students performing less well on this specific item (Figure 4). The difference of approximately 7.1 percentage points is statistically significant, as shown by the chi-squared test statistic of 168.54 (df = 1) and a p-value $< 2.2e-16$. This extremely low p-value provides strong evidence against the null hypothesis of equal proportions. Moreover, the 95% confidence interval for the difference in proportions ranges from -8.18% to -6.04%, suggesting that this gender gap is not due to random variation.

Figure 8: Results of the Z-test of the item 9, year 2017

```
        2-sample test for equality of proportions without continuity correction

data:  itemB out of totale
X-squared = 168.54, df = 1, p-value < 2.2e-16
alternative hypothesis: two.sided
95 percent confidence interval:
 -0.08175893 -0.06038475
sample estimates:
   prop 1    prop 2
0.2291036 0.3001755
```

These findings indicate that male students were significantly more likely to answer the item correctly. The consistent and substantial nature of the gap called for further investigation into the potential causes.
The results of the qualitative analysis highlighted response patterns among female students characterized by purely computational approaches devoid of interpretation, reflecting the formal proxy clause of the didactic contract (D'Amore, 2008). Some examples are presented in Table 1.

Table 2: Examples of female students' answers related to the formal proxy (translation by the author)

"I thought that in the 8 times table, sixteen appears twice."
"We know that division is the inverse operation of multiplication. Therefore, if you read the division backwards, you understand that 2 × 8 equals 16, which is why the division is correct."
"I calculated 8×16=120 and 120÷8=16."

The results revealed a correlation between the didactic contract and the incorrect responses provided by female students, which were characterized by a systematic reliance on calculations and multiplication tables as a problem-solving strategy, lacking critical reasoning. In some cases, the use of inverse operations led to justifications aimed at supporting procedural and interpretative errors, reflecting a complete delegation to algorithmic procedures, as described by the clause of formal proxy. The data were also compared with the incorrect responses of male students and the correct responses of females, showing a weaker association with formal proxy among males and a tendency among females, even in correct answers, to justify their responses through calculations. An example is provided below (Table 2).

Table 3: Some examples of answers of involved students (translated by the author)

Wrong answers of males	**Correct answers of females**
"I tried changing the divisor and multiplying."	*"I tried doing 8 ÷ 0.5 and got 16; I arrived at this result by carefully calculating how to reach 16."*
"Because I did 0.20."	*"I did 8 ÷ 0.5 and got 16."*

Therefore, the answers to the research questions emerge from the mixed-methods analysis. The quantitative analysis highlights the presence of a gender gap in specific mathematics items, based on data derived from standardized assessments. In contrast, the qualitative analysis provides possible interpretative frameworks for understanding the difficulties encountered by students when completing the questionnaire, particularly through the lens of the didactic contract. This perspective reveals how such a relationship may influence incorrect responses, especially among female students.

Conclusions

This study aligns with international research exploring gender disparities in mathematics performance and offers valuable insight into how specific standardized assessments items may reflect these differences. The analysis of INVALSI item 9 from the 2017 Grade 5 national test revealed gender gap, particularly in students' ability to reach the highest levels of proficiency. Quantitative analysis using the Z-test for two proportions showed that male students were more likely to select the correct response, while female students tended to choose the wrong answer. The tendency to give the wrong answer is higher for the female students of about 7%.
These results suggest that female students may rely more heavily on procedural strategies or calculation-based reasoning, often at the expense of interpreting contextual elements of a task. This behaviour is closely connected to the didactic contract, particularly the clause of formal proxy, which describes the delegation of meaning-making to purely formal mathematical operations. Qualitative data reinforce this interpretation, highlighting how females' explanations often reflect a strict adherence to operational procedures without engaging with the full context of the problem. The analysis revealed that while male students' incorrect responses appear to be more often linked to flawed reasoning detached from purely computational approaches, even the correct responses provided by female students tend to rely primarily on calculation rather than on deeper reasoning.
While the didactic contract is a mathematical construct, it does not exist in isolation from broader social influences such as gender norms and expectations. This study therefore contributes to a growing body of literature that seeks to understand the intersection between instructional dynamics and gender gaps. In particular, this research aims, on the one hand, to demonstrate the potential of standardized assessments

to highlight issues related to the gender gap, and, on the other hand, to explore possible connections with theoretical lenses from mathematics education. The goal is to promote a deeper understanding and inform teacher training, thereby contributing to the improvement of this critical situation. Indeed, the development of future research will aim to move in this direction, with the objective of identifying further connections of this kind and informing teacher training, to raise awareness among educators about these relationships and their implications for classroom practice.

Keywords: Gender Gap, Didactic Contract, Mathematics, Mixed Methods

Gender stereotypes and children's performance in primary schools

Piera Bello – Annalisa Cristini – Federica Origo

This paper investigates the impact of exposure to teacher stereotypes on student achievement in primary schools in Italy. We measure explicit stereotypes using item-based questions from the European Values Survey and elicit implicit gender bias using the Gender-Science Implicit Association Test.
By exploiting the random assignment of students to teachers with different levels of stereotypes, we show that the gender gap in math performance, defined as the difference between boys' and girls' scores on standardized tests, significantly increases in classes with math teachers with stronger gender stereotypes. Additional evidence suggests that this result is driven by girls who benefit from having teachers with unconventional stereotypes, that is a "girls-math" attitude.
In line with previous research, we do not find any effect of teacher stereotypes on student outcomes in reading.
We discuss some mechanisms and provide evidence on the role of the teacher's educational background in explaining our results.

Keywords: Gender stereotypes, education, human capital, implicit association

Understanding compositional effects: how classes' socio-economic and ethnic composition affect future educational trajectories

Costanzo Ranci – Andrea Parma – Francisco Ferraioli

Family backgrounds play a crucial role in determining children's educational outcomes, influencing not only school achievements (Invalsi 2024, OECD 2023) but also subsequent educational patterns. At the same time, also school contexts can contribute to the reproduction of educational inequalities by steering students along different educational paths (van Ewijk and Sleegers 2010; Langenkamp and Carbonaro 2018).
In particular, both family background and the characteristics of the schools attended in earlier grades can significantly affect the choice of academic tracks in upper secondary school, which in turn influences the likelihood of accessing tertiary education and impacts social mobility (Ben Ayed and Popeau 2009). Students from disadvantaged backgrounds are especially at risk of experiencing educational "Matthew effects": they are more likely to attend socially and academically segregated schools during the first cycle of education and to follow certain (often less prestigious) educational pathways in upper secondary school. The former tends to reinforce pre-existing inequalities and pre-determined trajectories aligning educational outcomes with students' socio-economic origins rather than promoting upward mobility.
Much attention has been placed on the effect of individual and family backgrounds while the research on the impact of compositions on future educational paths has been less investigated. However, school and classes compositions can play a role in shaping educational trajectories beyond the impact of individual features. By negatively hampering academic and professional trajectories and negatively impacting educational achievements, school segregation can be a driving force deepening educational inequalities.
While this line of research is more established in Anglo-Saxon contexts, it remains little developed in Continental and Southern Europe. This paper aims to fill this gap by focusing on Italy and in particular the city of Milano where recent socio-economic trends have reconfigured the composition of the students' population in the first cycle of the compulsory education (Fondazione ISMU 2024). Using microdata from

the standardized tests administered annually by Invalsi (National Institute for the Evaluation of the Educational System of Education and Training) to all students in Grade 8 (last year of the inclusive lower secondary education system) and applying regression techniques, we estimate the extent to which family backgrounds and socio-economic and ethnic compositions of the classes attended influences subsequent educational trajectories.

Specifically, we focus on the type of school attended in year 10 based on class composition of year 8's class and on the socio-economic and migration backgrounds of students' families. The socio-economic composition is measured using data on parental education attainments and occupations. Ethnic composition is analyzed considering migration backgrounds (and considering separately 1st and 2nd generations of foreign pupils). The analysis will pay a particular attention to the interaction between individual socio-economic backgrounds and the composition of students' intake of schools/classes attended to assess whether the influence of compositions varies based on individual attributes and to highlight which profiles are eventually are more strongly affected.

The main goal of this study, therefore, is considering whether and to what extent school segregation in lower schools tracks have an impact on the further educational career of students, controlling for individual characteristics and previous attainments. Our main research question is the following: To what extent does school composition explain the association between socio-economic position, ethnic background and educational attainments at individual level?

To address this research question, we will use a longitudinal analysis of the relationship between the composition of lower secondary schools and academic outcomes for students over an extended period of time. Using data from the INVALSI monitoring system for the lower secondary cycle and the upper secondary cycle, it will be possible to estimate the influence of schools composition in lower secondary school on educational choices and academic outcomes related to the upper secondary school, controlling for individual characteristics such as the educational attainments of parents, family income and nationality, and previous attainments. To measure educational outcomes, we will use standardised tests, as well as repetition of a school year and what kind of track students choose for upper secondary school.

Preliminary results show that socio-economic and migration backgrounds impact on the likelihood of choosing academic tracks (lyceums) in upper secondary schools. Students with parents with low socio-economic status and with migration backgrounds are more likely to choose vocational tracks or to drop-out from the upper secondary system. The trends remain significant even after controlling for previous achievements.

On top of that, a further effect can be detected: compositions of classes' intakes have a further additional impact on the likelihood to select a track over the others with a significant effect on choices of lyceums and on dropouts.

Socio-economic composition affects school choice more significantly than ethnic composition. The lower the share of low-ESCS students, the higher the chance that low-ESCS students are «missing» (training, dropouts, absents). Native students are more affected by ESCS composition than foreign students: in schools with higher proportion of low-ESCS students, the chance for native students to choose lyceums decreases more significantly that that of foreigners. IN such schools, as low-ESCS status, high-achieving students (G8) tend to choose more technical schools, while others are more likely not to take the G10 Invalsi test 2 years after their G8 test.

Ethnic composition shows little or no significant association with school choice once it is controlled by individual characteristics and socio-economic composition of classes attended. Students with migrant backgrounds are less likely to select lyceums but this seems to be driven by family backgrounds, previous achievements and socio-economic segregation.

These findings suggest that the Italian compulsory education system, though designed to be inclusive and comprehensive, contributes to shaping students' educational paths on top of the influence of family background alone.

Keywords: Composition, segregation, inequality, school choice

Intercultural Competence Profiles in Pre-Service Teachers: A Multidimensional Analysis

Giusi Antonia Toto – Guendalina Peconio – Alessandro De Santis - Pierpaolo Limone

Theoretical framework:
The growing cultural heterogeneity of Italian classrooms—highlighted by INVALSI (2023) data—makes intercultural competence a fundamental skill for educators. With over 900,000 students of non-Italian citizenship enrolled in schools (11.2% of the total), and more than 65% being second-generation immigrants, teachers must be equipped to navigate diverse cultural settings. Intercultural competence, as defined by Byram (1997) and Deardorff (2006), involves cognitive, affective, and behavioral dimensions essential for inclusive education.
Research objectives/questions:
This study investigates the latent structure of intercultural competence in a large sample of pre-service teachers and aims to identify recurring competence profiles. It also tests the reliability of the Cross-Cultural Competence Inventory (CCCI), providing empirical support for the design of differentiated teacher training programs.
Method of using INVALSI data:
The research builds upon the demographic context provided by INVALSI's 2023 national report, which emphasizes the significant presence of foreign-background students in Italian schools. These data serve as a foundation for the study's theoretical and pedagogical relevance, linking the empirical findings to national educational needs.
Results:
Based on responses from 1,098 pre-service teachers, a two-factor structure emerged: the first factor reflects openness and intercultural motivation, while the second is linked to defensiveness and emotional regulation. Cluster analyses (hierarchical, K-means, and Self-Organizing Maps) revealed six distinct participant profiles. High average scores were found in cultural adaptability, exploration, and goal orientation, while lower levels appeared in uncertainty tolerance and self-presentation.
Relevance to the Seminar:
This study directly aligns with the Seminar's themes by offering a robust empirical contribution on the intercultural readiness of future teachers. It also provides actionable tools—such as competence mapping and profile-based differentiation—to support inclusive school practices in increasingly diverse learning environments.

Keywords: intercultural competence, teacher training, factor analysis, sample segmentation

Evidencing the dynamics of educational inequalities: Attainment gaps in English secondary schools between 2018 and 2024

Nadir Zanini

Background and motivation
Attainment gaps are a key indicator of educational and societal inequalities. The analysis and reporting of educational attainment for different groups of students – and how it changes over time – is an important source of evidence for researchers, practitioners, and policymakers alike. Yet, the information available on attainment gaps is often lacking. Usually, it takes the form of simple and raw descriptive statistics that can be potentially misleading. In other instances, it relies on more in-depth investigations conducted as a one-off, which fail to provide systematic evidence to monitor the evolution of educational inequalities.
Research objectives
The aim of this presentation is twofold. From a methodological perspective, we will demonstrate that, despite their complexity, educational inequalities can be systematically measured and reported, if the appropriate data is made available. More specifically, we will outline an analytical research strategy that: i) accounts for the interplay between students' characteristics and other factors affecting their academic performance (including school-level effects); and ii) favours the analysis of over-time trends. From a more

substantive perspective, we will present a large amount of evidence on attainment gaps for different groups of secondary school students.

Methodology

We used a rich administrative linked micro-dataset, featuring a broad set of characteristics at individual- and school-level for all students in education at the age of 16 and 18 between 2018 and 2024. At student level, this includes standardised measures of educational attainment, prior schooling, socio-demographic characteristics (eg, gender, ethnicity, first language spoken at home, special educational needs), and socio-economic background. At school level, we used information on the type of school attended, its geographical location, and a measure of deprivation. A multivariate regression approach was employed to measure the impact of each characteristic, once other factors have been held fixed. To account for school-level effects, a multi-level regression approach was used. Each year's data was analysed separately. To interpret findings, a set of criteria was considered to flag over-time changes in attainment gaps that were considered 'notable'. Given the vast amount of findings available, a meta-regression approach was used to summarise the results and identify common patterns.

Results

We will present evidence on attainment gaps for students taking different education pathways in England. We will focus on the dynamics of educational inequalities between 2018 and 2024, drawing conclusions on the impact of the assessment arrangement put in place in 2020 and 2021 as a response to the pandemic and the return to a 'new normal'. We will show the mitigating (or sometime exacerbating) effects that school-level characteristics might have on the estimation of attainment gaps.

Relevance

By combining the rigour of in-depth studies with systematic monitoring, this study shows how it is possible to provide robust evidence on attainment gaps over time that takes into account individual- and school-level effects. Although using data on England, we will argue that this approach could be implemented also elsewhere.

Keywords: Multidimensional inequality, Attainment gaps, School effects, Over-time analysis

SESSION 1. TERRITORIAL DISPARITIES IN EDUCATION: A LONG-STANDING ISSUE ADDRESSABLE THROUGH NEW DATA AND TOOLS

ORGANIZER: UNIVERSITY OF MILAN BICOCCA
COORDINATOR: GIANLUCA ARGENTIN - ELISA MANZELLA
21ST NOVEMBER: 2.00 P.M. - 4.00 P.M. {AULA 4 – RESEARCH 15}

Educational Achievement and Territorial Inequalities in Italy: An Integrated Analysis of INVALSI and Open Government Data

Elisa Manzella

Inequalities in school learning represent one of the most persistent, complex, and relevant challenges facing the educational systems. Many national and international surveys on student achievement, measured through standardized assessments, have clearly highlighted the existence of marked educational inequalities.

This contribution focuses on the Italian context, where previous studies have shown that students' socio-economic background significantly influences learning gaps (Argentin et al., 2017a; Giancola & Salmieri, 2020). Other researches have focused on the role of the school context and the impact of the different socio-economic composition of schools across different territories (Argentin et al., 2017b).

Another factor that may affect learning inequalities is the territorial context. The Italian educational system is characterized by marked territorial disparities, with a significant disadvantage for students residing in the southern regions compared to their peers in Central and Northern Italy. Extensive empirical evidence gathered over the past decades confirms the existence of this structural and persistent gap between the different geographical areas of the country (Argentin et al., 2017b; Asso et al., 2015; Bratti et al., 2007; Foresti & Pennisi, 2007; Lo Cicero & Giancola, 2025; Sestito & Bonvini, 2021). This gap emerges as early as primary school and tends to widen throughout students' educational trajectories. This phenomenon affects the possibility of access to quality educational path and adequate employment opportunities, thus contributing to the reproduction of social and territorial inequalities.

Several studies have highlighted the central role of socio-economic and cultural factors, as well as elements related to geographical and territorial characteristics, in shaping students' competence levels, as measured through standardized tools such as the INVALSI assessments (Argentin et al., 2017b; Sestito & Bonvini, 2021; Sacco & Falzetti, 2021). However, a more detailed and updated understanding is still needed of the mechanisms through which these factors operate, interact, and combine in producing educational inequality. There is a need for analyses capable of integrating data from multiple sources, in order to better capture the multidimensional nature of educational disparities (Sacco & Falzetti, 2021).

In light of these considerations, thi present study aims to identify the main explanatory factors underlying the territorial gaps in school achievement, focusing on the results obtained by Italian students in the INVALSI standardized tests in Italian and Mathematics, administered in both the fifth grade of primary school (Grade 5) and the eighth grade of lower secondary school (Grade 8). Specifically, the analysis investigates the contribution provided by different categories of variables, including the individual and family characteristics of the student, the territorial and geographical context at the municipal and local labour systems levels, as well as data on the cultural vocation of each territory.

The analysis was conducted by integrating data provided by INVALSI, related to students in the fifth grade of primary school (Grade 5) and the eighth grade of lower secondary school (Grade 8), with a rich territorial information base developed within the research project "OPEN GOVERNMENT DATA. Conoscere la società attraverso i dati della Pubblica Amministrazione" – Finanziamento MUR 2023-2027 – Nota 15659 del 28/12/2022), promoted and funded by the Department of Excellence in Sociology and Social Research at the University of Milano-Bicocca. This integration allowed the inclusion of a wide range of explanatory variables, organized across different levels: individual and family characteristics at the student level, school and territorial contextual features, including their economic and cultural dimensions.

The methodology employed is based on multivariate linear regression models, with the aim of isolating the specific contribution of the various factors in determining student achievement levels, measured through the standardized INVALSI scores. The analysis is conducted separately for the two assessed subject areas (Italian and Mathematics), in order to highlight possible differences in territorial learning patterns. In

addition, particular attention is given to the dynamics of skill development (learning delta) between Grade 5 and Grade 8, which helps reveal differences in competence growth rates depending on geographical area. The results confirm the existence of significant territorial differences in average competence levels. In particular, students in Northern regions tend to obtain higher INVALSI scores compared to their peers in the South. These differences can be partly explained by individual and family-level variables, as well as by territorial factors. The progressive inclusion of contextual variables in the model accounts for an increasing share of the variance in scores, reducing—though not eliminating—the geographical gap. These territorial differences persist even after controlling for individual and family characteristics, suggesting a significant role played by broader territorial factors.
Competence growth between Grade 5 and Grade 8 follows a similar pattern. Data show that students in the South not only start from lower levels, but also experience slower growth compared to students in the North. Indeed, the average learning delta is significantly lower in the Southern area. Even after controlling for all available individual, family, and territorial variables, the growth gap remains statistically significant, confirming the persistence of disadvantage in the South.
The findings suggest the existence of structural and systemic mechanisms that hinder the functioning of the school system in some areas of the country more than in others, beyond the factors measured and included in the analysis. These results have important implications both from a scientific and a policy perspective. On the one hand, the study confirms the need for multidimensional and territorially sensitive approaches to the analysis of educational inequalities. On the other hand, it highlights the urgent need for targeted interventions aimed at reducing learning gaps, which cannot be limited to acting on individual students or schools alone. These interventions must include integrated local development policies, also aimed at enhancing cultural opportunities and supporting the most fragile territories.
This study contributes to the debate on educational inequalities in Italy by providing a solid and up-to-date empirical basis for the design of territorially targeted educational policies that aim to reduce territorial gaps and promote greater equity in the school system.

Keywords: Territorial inequalities, student achievement, educational opportunity, data integration

The use of georeferenced data to analyze territorial disparities in Education

Michele Marsili – Fiammetta Noccioli – Patrizia Falzetti

Introduction

The Italian school system is characterized by a marked territorial heterogeneity, which is reflected in the contrast between the North and the South of the country. INVALSI data also confirm these disparities, showing significant differences in educational outcomes between different geographical areas. As discussed by Costanzo & De Simoni (2017), territorial inequalities in italian and mathematics skills emerge from the early years of primary school. The *Rapporto sulla scuola media 2021* by Fondazione Agnelli also points out that these inequalities tend to become more pronounced in secondary school, underlining how the latter and the territorial gaps become more pronounced than in primary school. The identification of learning gaps, however, is particularly complex when the analysis shifts from the macro-area level to that of smaller areas, which can, however, hide significant differences as, for example, within urban contexts such as Turin and Milan (Falzetti et al.,2025).
Regardless of the territorial division adopted, socio-economic factors maintain a decisive role in generating and widening the gap. The PISA 2022 data in mathematics indicate that in Italy about 13% of the variability in scores is explained by the SES (Socio-Economic Status) Index of students. However, there are differences in the relationship between geographical macro-areas and socio-economic level. For example, in the South, socio-economic level explains 15% of the variability against 8% observed in the Centre.
According to the INVALSI Report 2024, the average advantage of male and female students with a more favourable socio-economic level is 7.2 points. To this, a further 9.4 points must be added if the school receives on average students with higher ESCS (Economic, Social and Cultural Status) levels. Therefore, not only do individual factors affect the territorial differences but a not insignificant part of the variability of the phenomenon can be attributed to differences in the characteristics of the schools.

Research object

Firstly, the aim of the study is to analyze how the relationship between socio-economic level and learning achievement across territory that overcomes the concept of administrative unit. Secondly, having highlighted the strong territorial variability in mathematics scores, we wanted to investigate whether this trend is also confirmed for italian. Spatial analysis techniques may be a valid solution in this context to investigate spatial dynamics and to explore the factors of inequality. Lastly, particular attention was paid to the choice of the most suitable clustering methodology for the study context. Indeed, the use of these techniques makes it possible to identify, among other things, schools that are more sensitive to socio-economic conditions.

Data

The study is based on a dataset constructed from the latest available INVALSI data on students in grade 8. Specifically, Comprehensive Institutes belonging to all Italian regions were included, with the exception of the autonomous provinces of Bolzano and Valle D'Aosta due to lack of data.For reliable results, only institutions with a participation rate in the National Surveys greater than 50 percent, for both Italian and mathematics tests, were considered. From the individual data of students who participated in the National Surveys, averages of WLEs (Weighted Likelihood Estimates) in both subjects were calculated at the school level. Contextual information was also considered, based on the relevance and completeness of the data, such as ESCS, migration background, gender and regularity in studies. Finally, each institution was associated with its spatial coordinates. Georeferencing was done from the addresses of the main buildings (school secretary) of schools available in the INVALSI database.

Methods

Through spatial coordinates it was possible to calculate the distance matrix that allowed us to verify the presence of spatial autocorrelation, in our context, if geographically closer schools present more similar values in the INVALSI results. Moran's I-index, a classic tool in spatial statistics, was used for this analysis. The results showed a significant level of autocorrelation, justifying the adoption of such a model. Therefore, spatial GWR (Geographically weighted regression) models were applied using the WLE of grade 8 as the dependent variable for both Italian and Mathematics. An adaptive kernel was also applied to account for the fact that schools are unevenly distributed. This is particularly useful in contexts characterized by strong territorial discontinuity, such as Italy, where high-density urban areas and more isolated rural areas coexist. These models were compared with OLS (Ordinary Least Squares) models to verify the actual improvement in the results. Finally, different clustering methodologies were tested such as Getis-Ord Gi*, which measures the degree of spatial concentration of an attribute in an area with respect to surrounding areas, highlighting significantly high (hot spot) and low (cold spot) values, in order to obtain an effective and granular representation of the phenomenon.

Results

In this study, GWR was applied in combination with territorial clustering techniques for better representation of territorial disparities, useful for targeting more effective educational policies. This approach made it possible to identify groups of schools in which socio-economic conditions have a greater impact on INVALSI test. The results show that the regression coefficients vary spatially, indicating that the influence of socio-economic factors on INVALSI test results at grade 8 is not homogeneous.

The adoption of a spatial approach made it possible not only to map critical areas but also to identify potentially more vulnerable institutions. This method could prove to be a useful synthesis tool and relevant for policy purposes such as a more efficient planning of educational interventions by adapting to the specificity of local contexts. The methodology used could be extended to other school levels and integrated with further context variables, thus contributing to a deeper understanding of educational inequalities in Italy.

Keywords: Territorial disparities**,** spatial analysis, Clustering, INVALSI data

Clouded minds: the impact of air pollution on students' cognitive performance

Simone Ferro – Elena Meschi – Caterina Pavese

Introduction

Air pollution poses a major environmental and public health challenge, with growing evidence of its widespread impact not only on physical health outcomes but also on human behavior, cognition, and economic performance. According to the World Health Organization, ambient air pollution is responsible for over 4 million premature deaths globally every year. While its medical consequences—particularly regarding respiratory, cardiovascular, and neurological conditions—have been extensively studied, recent research has begun to uncover how environmental quality can also affect cognitive functioning, especially among children. This paper contributes to this emerging field by examining the short-term effects of air pollution exposure on students' academic performance in Italy, focusing specifically on primary school pupils, an age group often overlooked in previous analyses. Italy offers an important case study given its frequent violations of air quality standards, especially in urban and industrial areas of the northern regions.

Research Object, Aims, and Hypotheses

Our primary research question is whether temporary exposure to fine particulate matter (PM2.5) during exam days affects students' performance in standardized school tests. Specifically, we aim to estimate the short-run causal impact of pollution on cognitive outcomes as captured by INVALSI standardized assessments in Italian, Mathematics, and English among second- and fifth-grade students. We hypothesize that higher PM2.5 concentrations on test days negatively affect cognitive functioning and academic performance, with stronger effects expected on tasks that require complex reasoning and higher-order thinking, rather than basic knowledge recall.

Data Sources

The study uses a comprehensive dataset constructed by merging three main sources:

1. **INVALSI data** (2012/13–2022/23): This dataset includes standardized test scores for approximately 22 million student-subject observations across ten cohorts in primary school. The assessments are conducted on fixed dates nationwide, allowing precise linkage to daily environmental data. Students are tested in different subjects on different days, which provides multiple data points per student, enabling within-individual comparisons.
2. **Air pollution and weather data**: PM2.5 concentrations are derived from the Copernicus Atmosphere Monitoring Service (CAMS) at high temporal and spatial resolution. Weather variables—temperature, wind, precipitation—are taken from ERA5 reanalysis data. We also include the height of the planetary boundary layer (PBLH) as a variable influencing pollution dispersion, which is used in our instrumental variable strategy.
3. **School infrastructure data** (MIUR): We incorporate municipality level information on whether school buildings are equipped with air conditioning or air filtration systems, and whether they are located in environmentally or socially disadvantaged contexts (e.g., near polluting industries, in high-traffic zones, or degraded urban areas).

Methodology and Identification Strategy

Our empirical strategy relies on panel data methods exploiting within-student and within-municipality variation in pollution across test dates. The core regression model includes an extensive set of fixed effects: student, year, grade, subject, municipality, day-of-week, and interactions between grade, subject, and year. This rich structure controls for unobservable factors at both the individual and local levels. We further adjust for contemporaneous weather conditions.

To strengthen causal identification and address potential endogeneity, we employ an instrumental variable (IV) approach using the inverse of the planetary boundary layer height (PBLH) as an instrument for PM2.5. This meteorological variable, which affects the vertical dispersion of pollutants, is plausibly exogenous to students' academic performance after controlling for temperature and precipitation.

Main Results

Our results provide robust evidence of a statistically significant negative effect of PM2.5 exposure on student performance. Specifically, we find that a one microgram per cubic meter ($\mu g/m^3$) increase in PM2.5 concentration on test day leads to a reduction of 0.14 points in test scores, corresponding to a 0.33% decline relative to the standard deviation of the test scores. These results are stable across multiple specifications and remain significant when controlling for time-varying confounders and individual heterogeneity.

We explore several dimensions of heterogeneity and mechanisms:

- **Cognitive domain**: The effect of pollution is stronger for reasoning-based questions than for knowledge-based ones. This suggests that air pollution particularly impairs cognitive processes related to attention, logic, and problem-solving, which are more demanding and vulnerable to environmental stressors.
- **Question format**: Multiple-choice questions are more affected than open-ended questions. While this may seem counterintuitive at first, it aligns with research on the "paradox of choice" and cognitive overload: the need to evaluate and discard incorrect options under polluted conditions can exacerbate stress and reduce performance.
- **Age and ability**: Younger students (second grade) are more affected than older ones (fifth grade), supporting the view that younger children are biologically more vulnerable to air pollution due to their developing neurological systems and higher pollutant intake rates. Moreover, students in the top quartiles of initial ability (measured in second grade) suffer disproportionately, possibly because they are more engaged in reasoning-heavy test strategies.
- **Environmental context**: The detrimental effect is amplified in schools located in disadvantaged or polluted areas—such as those close to industrial activity, traffic congestion, or environmental degradation—highlighting environmental inequality in educational outcomes.
- **Emotional responses**: We also analyze student responses to a test-day survey on emotional states. We find that pollution exacerbates the negative effect on students who report high levels of worry, anxiety, or insecurity, suggesting that air quality interacts with emotional regulation and test-related stress, compounding its negative impact on performance.

Role of Infrastructure and Mitigation

A key policy-relevant finding is that the presence of air filtration or air conditioning systems significantly attenuates the effect of pollution. In municipalities where schools are equipped with air ventilation systems, the magnitude of the negative effect is reduced by nearly 50%. This finding supports the idea that relatively low-cost interventions in school infrastructure can effectively shield students from environmental risks and reduce inequality in academic opportunities.

Conclusion

This study provides novel and compelling evidence that short-term exposure to fine particulate matter (PM2.5) can impair the cognitive performance of children in a standardized testing context. These effects are stronger in reasoning-intensive tasks, among younger and higher-achieving students, and in polluted or disadvantaged environments. Moreover, we highlight the role of emotional vulnerability and school infrastructure as factors that moderate these effects.

The implications are twofold. First, environmental quality should be considered an integral part of educational policy, particularly in urban and industrialized areas. Second, improving school infrastructure, especially through better ventilation and filtration systems, offers a concrete and feasible avenue to protect cognitive development and promote equity in educational achievement.

Keywords: Air pollution, test scores, cognitive ability, PM2.5

Systemic Learning Loss or Territorial Disparities? The Impact of the Pandemic on Student Achievement and School Value-Added in Italy

Giulia Assirelli – Salvatore Vassallo

Territorial disparities represent one of the most persistent and problematic dimensions of the Italian education system. Numerous studies have documented how learning inequalities across different areas of the country are both wide and structural, often intersecting with broader social and educational inequalities. The COVID-19 crisis has raised further concerns about the resilience of the system and the possible widening of such disparities, but it remains an open question whether the deterioration in student performance has been uniform or territorially differentiated.

This contribution aims to analyze the evolution of student achievement in Italy over the period 2017/18 to 2021/22, with a specific focus on the impact of the COVID-19 pandemic. The research addresses two main questions: (i) whether the decline in student learning observed in the post-pandemic years was generalized or whether it led to an intensification of pre-existing territorial disparities; (ii) how the value-added produced by schools evolved over time across different geographical contexts.

The analysis draws on INVALSI test scores in Italian and mathematics for grade 8 students, collected from over 7,200 schools for which complete data are available for all four school years considered. The comparability of scores over time is ensured by anchoring the results of later years to the 2017/18 baseline. At the school level, a value-added indicator was also computed, based on models that control for individual and school-level student characteristics. Preliminary exploratory analyses also examine associations between performance trends and territorial as well as structural characteristics of schools.

Findings show that the overall decline in student achievement affected the entire country, but with varying intensity: in southern and insular regions, the decline was more pronounced, particularly in mathematics. However, the analysis of school value-added reveals a more nuanced picture: some areas in the South and Islands show relative improvement in the national distribution, whereas schools in the Center and North-West exhibit a gradual decline. Furthermore, schools that had recorded higher average performance levels before the pandemic tended to experience greater losses. The analyses also consider how certain school- and territory-level characteristics may have contributed to mediating these dynamics.

This contribution offers new empirical insights into the evolution of territorial disparities during a critical phase in the recent history of Italian education. Demonstrating that the pandemic's impact was not uniform—and that some schools were able to activate more effective responses than others—provides valuable input for rethinking equity-oriented territorial policies and for guiding public action toward strengthening the transformative capacity of the most fragile educational contexts.

Keywords: Territorial disparities, educational achievement, value-added, pandemic

Inequality of Educational Opportunities by Social Background: Geographical Variations and Contextual Drivers in Italy

Moris Triventi – Emanuele Fedeli – Andrea Pietrolucci – Nathalie Vigna

Introduction and theoretical framework

Despite the formal centralization of many education systems, inequalities in student outcomes and educational pathways vary consistently between regions within countries, mirroring deep-rooted spatial divides (Betthäuser et al., 2021; Rodríguez-Pose & Tselios, 2011). Italy exemplifies this pattern: stark regional disparities in both student learning and access to educational opportunities coexist within a single, nationally regulated school system. While much attention has been paid to differences in average performance across regions, fewer studies have focused on how the strength of social background effects on education attainment varies across local contexts and how such variation relates to contextual characteristics.

Extensive research in the field of educational stratification has shown that family background is a key factor in determining students' educational trajectories (Blossfeld et al., 2016; Breen et al., 2009). pecifically, the transition to upper secondary education, where choices between academic and vocational pathways are made, represents a pivotal point in the intergenerational transmission of advantage (Jackson, 2013; Triventi et al., 2021). The dominant theoretical approach to explaining such inequality is the theory of rational action (Breen & Goldthorpe, 1997), which highlights the role of cost-benefit calculations, risk aversion and family expectations.

At the same time, a growing body of work highlights the geographical heterogeneity of social mobility, showing how individuals' life chances depend not only on their family of origin but also on the places where they grow up (Betthäuser et al., 2021; Breen & In, 2024; Chetty et al., 2014; Rodríguez-Pose & Tselios, 2011). These studies often rely on large-scale administrative datasets and point to significant territorial differences in outcomes such as income, education, and occupational attainment. However, they usually focus on outcomes rather than mechanisms, and often lack detailed information on early academic performance or proxies of individual ability (Chetty et al., 2014; Rodríguez-Pose & Tselios, 2011). Conversely, survey-based studies with rich individual-level information are limited by smaller samples and coarser geographic units, making it difficult to capture the fine-grained spatial variation that characterizes many educational systems (Betthäuser et al., 2021; Breen & In, 2024). This disconnect has left a gap in the literature regarding how regional contexts interact with social background to shape key educational transitions, and through which processes.

Research questions and hypotheses

This study contributes to bridging these fields by investigating the following questions:

1. To what extent does the relationship between parental education and academic track enrolment vary across Italy's local labour market areas (LMAs)?
2. What contextual factors—economic, institutional, or educational—help explain this territorial variation in the inequality of educational opportunity (IEOp)?

The central hypothesis is that territorial opportunity structures condition the strength of social background effects on students' educational pathways. These local conditions may either reinforce or mitigate the mechanisms of social reproduction, depending on families' perceptions of available options and risks. Building on theories of rational action and the geography of opportunity, the study formulates six hypotheses:

- **H1:** Social inequality in academic track enrolment is lower in economically prosperous areas.
- **H2:** Such inequality is higher in areas with greater income inequality.
- **H3:** areas relying more on the industrial economic sector (such as the north-east Italian regions) exhibit higher inequality due to the attractiveness of vocational pathways.
- **H4:** Higher youth unemployment is associated with stronger social background effects due to increased risk aversion among lower-SES families.
- **H5:** Higher school quality in lower secondary education mitigates social inequality in track enrolment.
- **H6:** Greater local supply of academic schools (lyceums) reduces inequality by increasing access for all students.

Data and analytical strategy

The analysis uses a unique combination of longitudinal administrative microdata from the Italian national student assessment system (INVALSI-SNV) and rich macro-level administrative data for over 540 LMAs. LMAs are defined based on commuting patterns and represent the appropriate spatial unit to capture both school catchments and labour market conditions. Two student cohorts (born in 2004 and 2005) are followed from Grade 5 through Grade 10, resulting in a dataset of over 560,000 students.

A two-step multilevel strategy is used to estimate and explain variation in social background effects. In the first step, we compute LMA-specific marginal odds ratios capturing the association between parental education and enrolment in the academic track (lyceum), both gross and net of academic performance. In the second step, these measures are regressed on contextual variables—including youth unemployment, industrial employment share, income levels, school availability, and secondary school quality—using a model that accounts for the uncertainty in the first-step estimates.

Main findings

The findings reveal substantial spatial heterogeneity in educational inequality. As reported in Figure 1, the strength of social background effects on students' enrolment decision varies significantly across LMAs, both when looking at the total effect (left panel) and when controlling for students' prior achievement (right panel). This suggests that students' family origins matter more—or less—depending on where they live, despite the centralized nature of the education system.

In the second step of the analysis, we tested the association between the level of inequality in each LMAs and several contextual predictors. Figure 2 shows the results from such models. Several contextual variables are found to be robust predictors of the geographical variation in enrolment inequality. Youth unemployment is strongly and positively associated with both gross and net inequality, suggesting that families in high-risk labour markets may make more cautious and protective educational decisions, particularly advantaged families who can better shield their children. Moreover, industrial employment share correlates positively with inequality, indicating that vocational pathways embedded in prosperous manufacture-based local labour markets may be more attractive to low-SES families, thereby reinforcing inequality. Finally, local supply of academic schools is associated with lower net inequality, suggesting that a broader range of accessible options can help level the playing field for equally performing students from different social origins.

Interestingly, average income per capita and income inequality (GINI index) show weaker and more ambiguous associations with social background effects. These results suggest that opportunity structures and institutional arrangements may be more influential than general economic affluence in shaping families' educational strategies.

Figure 1. Choropleth map of marginal odds ratios (MORs) of the total (left panel) and direct (right panel) effect of parental education on academic track enrolment across LMAs

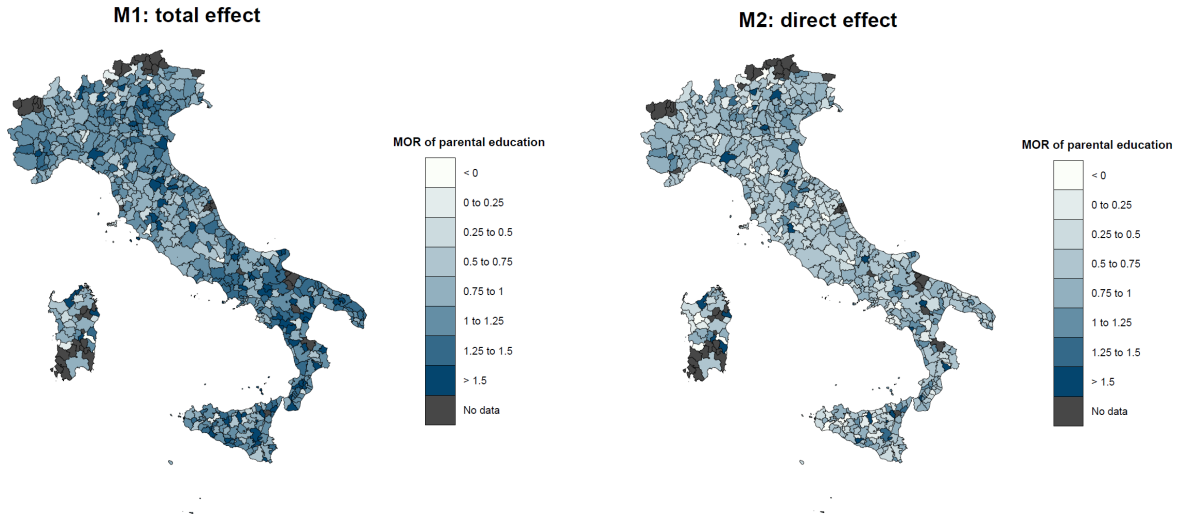
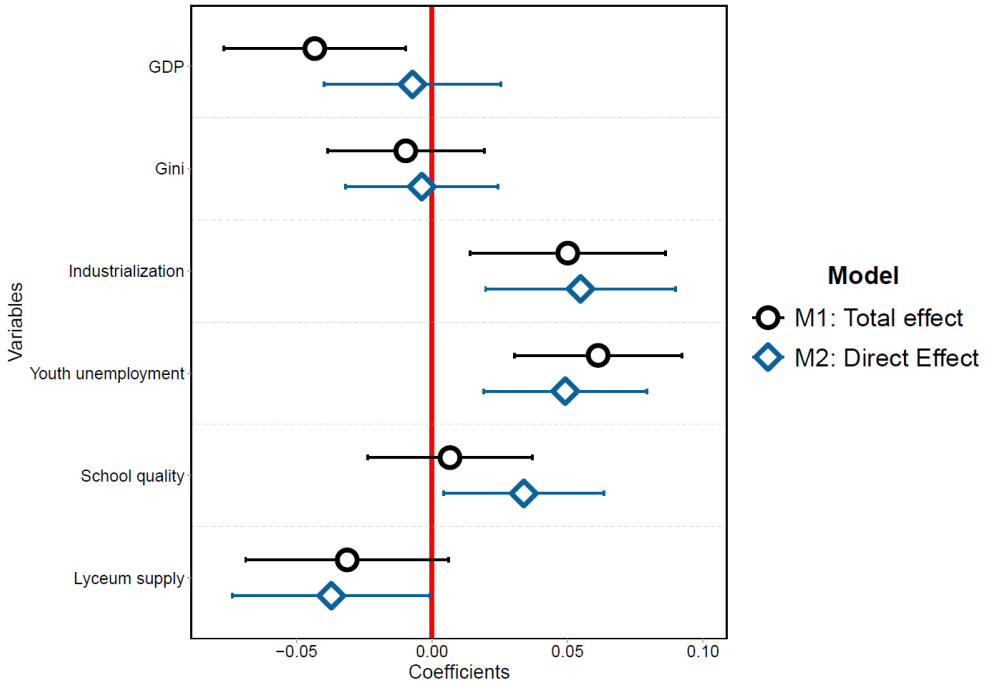


Figure 2. Two-step multilevel analysis: effects of contextual variables on MORs of parental education (total and direct effect) on the probability of academic track enrolment (coefficients and 95% confidence intervals)



Conclusion

These findings carry significant implications for the study of territorial disparities in education. They demonstrate that inequalities in access to academic tracks are not uniform, even under a common institutional framework. The spatial variation uncovered here adds a new dimension to the geography of inequality and calls for greater territorial sensitivity in stratification research. Moreover, our analyses underscore the potential of high-resolution longitudinal data—such as those produced by INVALSI—to illuminate mechanisms of inequality that operate below the national level. Combining these with context indicators enables a much more detailed mapping of educational opportunity structures. Finally, the study illustrates the importance of meso-level policy instruments—such as improving access to lyceums, reducing youth unemployment, or reforming guidance programs—that can be deployed at the regional or municipal level to mitigate the effects of family background.

Overall, this study contributes to a more nuanced and spatially aware understanding of educational inequality in Italy. It highlights how local opportunity structures—rather than national policies alone—mediate the relationship between social origin and educational decisions. From a policy perspective, the findings call for place-sensitive interventions that account for both the structural constraints and the strategic behaviour of families.

Keywords: Inequality of educational opportunity, tracking, geographical inequality, contextual predictors

Peer Socioeconomic Effects on Academic Track Enrollment

Emanuele Fedeli – Moris Triventi

Introduction

This contribution is part of the scientific debate aimed at investigating how belonging to learning environments with different socioeconomic profiles affects educational opportunities and selection processes in school pathways. International literature has extensively discussed the existence of contextual effects linked to the social composition of classrooms—the so-called peer effects (Marks 2015; Nash 2003; Triventi, Barone, and Facchini 2021; Van Ewijk and Sleegers 2010)—and their role in either fostering or discouraging the aspiration to undertake more selective academic tracks. Previous studies have shown that the presence of classmates from more advantaged backgrounds can exert a positive influence on individual motivation, expectations, and self-efficacy, while at the same time generating mechanisms of social comparison and a reduction in self-esteem among students from families with lower socioeconomic status (Van Ewijk and Sleegers 2010). Moreover, the learning environment and teachers' expectations tend to be shaped by the average composition of the class. This work aims to analyze these mechanisms in the Italian context, characterized by a tracking system starting at the age of 14, with the goal of assessing to what extent family background and classroom context contribute to students' decisions to enroll in different upper secondary school pathways and to what extent a territorial divide comes out.

Research Objectives

The research focuses on the transition from lower secondary to upper secondary school, a phase in which students are required to choose among academic, technical, or vocational tracks—a decision that has long-term effects on subsequent educational and occupational trajectories. The main objectives revolve around two research questions:

1. How does individual socioeconomic status influence enrollment in academic (lyceum) tracks compared to less academic alternatives?

2. To what extent does the socioeconomic composition of the originating class (measured in the final year of lower secondary school), independently of individual characteristics and prior performance, affect the probability of undertaking a more prestigious educational path?

In light of these questions, we hypothesize that exposure to classmates from more advantaged backgrounds is associated with a higher propensity to choose the lyceum track, both through the raising of individual expectations and as a result of normative pressure that values pathways considered to have higher social prestige (Davis 1966; Festinger 1954). At the same time, we consider the hypothesis that the interaction between individual SES and classroom SES composition may amplify or attenuate these dynamics.

3. The moderating effect between individual SES and classroom SES may vary depending on the economic deprivation of the area of residence.

Data

The analysis relies on INVALSI-SNV data concerning the cohorts who attended the fifth grade of primary school in 2012, 2013, and 2014. These data provide information on a sample of approximately one and a half million students belonging to two different cohorts, who are then followed in the last year of lower secondary school and in the second year of upper secondary school. For each student, detailed indicators are available on socioeconomic status, reconstructed by combining parental education and occupation with the number of books in the household, together with standardized test scores in mathematics and reading in the second year of upper secondary school. The main outcome variable is the type of pathway chosen at the end of lower secondary education: lyceum vs other tracks.

Method

From a methodological perspective, we adopt linear regression models with school fixed effects, in order to neutralize potential confounding factors at the school level, such as guidance policies or overall reputation. This strategy allows us to isolate variation in socioeconomic composition among classes within the same schools. A crucial aspect of the approach involves the implementation of a sorting test (Contini 2013; Fedeli and Triventi 2024), aimed at identifying and selecting only those schools that do not systematically distribute students among classes based on individual characteristics or prior academic performance. The regression models include interactions between individual SES and the average classroom SES composition, while also controlling for gender, migration background, prior academic performance, geographic area,

cohort, and measures of individual and classroom competence. To reduce the risk of bias arising from attrition over time, we employ Inverse Probability Weighting techniques (Seaman and White 2013).

Results

Preliminary results indicate that individual socioeconomic status is a very strong predictor of the probability of enrolling in the lyceum track, being associated with an estimated increase of about eight percentage points in the likelihood of this choice compared to peers with less advantaged backgrounds. The average classroom composition exerts a statistically significant, though smaller, influence, confirming that the social environment of peers contributes to shaping educational decisions. However, the interaction between individual status and classroom composition appears less intense than hypothesized: the contextual effect does not seem to systematically increase according to the student's family background. Furthermore, the robustness of these findings diminishes when controls for cognitive skills measured by standardized tests are introduced, suggesting that part of the classroom composition effect may reflect pre-existing differences in academic performance.

An ongoing phase of the analysis involves the integration of more detailed information on students' areas of residence. Specifically, an indicator of exposure to public services will be developed by combining the presence of hospitals, childcare centers, libraries, museums, schools, transportation hubs, and other infrastructure, each weighted by the distance from the place of residence. This indicator will subsequently be divided into three access categories (low, medium, and high). Preliminary results show that exposure to a low availability of public services is associated with a stronger influence of socioeconomic background on the probability of enrolling in academic tracks and with a more pronounced effect of classroom social composition. These findings suggest that territorial inequalities and conditions of material deprivation contribute to amplifying the mechanisms that reproduce educational inequalities.

Keywords: Peers, classroom, socio-economic background

SESSION 2. EXPLORING EDUCATIONAL POVERTY: CONCEPTUAL AND METHODOLOGICAL APPROACHES THROUGH THE INTEGRATION OF OFFICIAL STATISTICAL DATA SOURCES

ORGANIZER: ISTAT

COORDINATOR: BARBARA BALDAZZI

21ST NOVEMBER: 4.15 P.M. - 6.15 P.M. {AULA MAGNA – RESEARCH 16}

Measuring Educational Poverty: An Empirical Comparison of Two Indices Using Individual-Level Data

Giulia Biazzi – Francesca Milzani

Introduction

In recent decades, the value of education—understood primarily as formal schooling—has been strongly reaffirmed at the international level as a fundamental right and an essential condition for human development, active citizenship, and personal fulfillment. In the key strategic documents of the United Nations and UNESCO—from the UN Convention on the Rights of the Child (1989), to the Faure Report (1972), to the Delors Report (1996)—education is presented not only as a driver of economic growth, but also as a means to "live together," strengthen social cohesion, and promote individual freedom. Similarly, the European Union has placed education at the core of its long-term strategies, from Europe 2020 to the 2030 Agenda for Sustainable Development, highlighting its role in fostering social inclusion and sustainable development.

However, within this shared framework, the European debate on educational poverty has developed largely around a narrow conception of education, focused on school outcomes, early school leaving, and basic skills levels (Education and Training Monitor, European Commission, 2015).

In Italy, since 2014, a broader and more multidimensional conception of educational poverty has emerged, strongly promoted by Save the Children Italia and later developed by ISTAT. In this perspective, the concept does not coincide merely with the failure to attain formal qualifications or minimum standardized skills, but refers instead to the deprivation of educational and formative opportunities—both formal and informal—that enable children and adolescents to fully develop their abilities, aspirations, and relationships (Save the Children Italia, 2014; Quattrociocchi, 2018; Pratesi, 2020, 2022; ISTAT, 2024). This view is explicitly inspired by the capabilities approach (Sen, 1999; Nussbaum, 2011), developed within the field of multidimensional poverty studies, which emphasizes the importance of providing individuals with the real opportunities to develop their potential and lead a fulfilling life. From this perspective, educational poverty represents an inability to fully participate in society.

In the Italian context, both a narrower and a broader view of educational poverty coexist, contributing to definitional ambiguity and operational uncertainty. The strength of the approach proposed by Save the Children and ISTAT lies in the attempt to capture the complexity of the phenomenon by taking into account multiple factors—individual, familial, school-related, and territorial—in the measurement of educational poverty. The evolution of the indices developed by these two institutions between 2014 and 2024 highlights an increasing effort toward both theoretical and operational refinement, shifting from an initial focus on the provision and use of educational services to a more sophisticated measurement approach that includes cognitive and non-cognitive outcomes, as well as life context variables.

Nonetheless, the absence of a fully shared theoretical framework capable of clearly and systematically defining the very concept of educational poverty remains. In the absence of such a definition, the construction of measurement tools becomes more complex. The choice of which dimensions to consider, which indicators to include, and which thresholds to adopt can vary significantly depending on the underlying conceptual approach. This directly affects how the phenomenon is delineated and represented, also influencing which groups are identified as priorities and how public intervention is oriented.

Aim, Objectives and Research Hypothesis

This paper aims to analyze how different approaches to defining and measuring educational poverty influence the empirical representation of the phenomenon and, consequently, the priorities of public policy interventions. In particular, it reflects on the implications of including a varying number of dimensions and

indicators, as well as on the opportunity to combine contextual factors and outcomes within a single composite index.

The analysis is part of a broader critical review of the evolution of educational poverty measurement strategies in Italy and of the national and international theoretical debate. Building on this framework, the paper presents a comparative empirical exercise, applying two versions of composite educational poverty indices — Save the Children 2014 and ISTAT 2018 — to the same dataset. Although a more advanced measurement proposal is now available (EPI ISTAT 2024), the decision to use these "previous" indices is based on two main considerations: on the one hand, their actual availability at the time of data collection; on the other, the scientific interest in highlighting — through a direct comparison using individual-level data — the potential implications of different measurement approaches for the representation of the phenomenon.

The hypothesis is that the application of different indices, even when based on the same dataset, generates divergent representations of educational poverty. This raises important questions about the relationship between theoretical frameworks, measurement operationalization, and the informational potential of the various indices.

Data Used

The empirical analysis was conducted using a dataset collected within the framework of an educational project implemented in three territorial areas of the province of Brescia – Val Sabbia, Val Trompia, and Bassa Bresciana Centrale. In total, the data were gathered from 28 schools, including 23 comprehensive institutes and 5 upper secondary schools. Approximately 730 questionnaires were administered to students, with the aim of collecting indicators related to educational poverty before and after the project. The questionnaires were designed to cover multiple dimensions of educational poverty as identified at the time by the available indices (Save the Children 2014 and ISTAT 2018). While the dataset does not fully include all indicators originally proposed by each index, it contains variables that are sufficiently comparable to allow for a reasoned reconstruction of both measurements, capturing their intended dimensions. The dataset is individual-level and locally sourced, collected directly in schools rather than derived from administrative or aggregated statistical sources.

Method or Approach

The adopted methodology involves a systematic comparison between two composite indices of educational poverty – Save the Children 2014 and ISTAT 2018 – which, despite differing in structure and complexity, share several analytical dimensions and include partially overlapping indicators. The empirical exercise consists of applying both indices to the same dataset and analyzing the differences in how subjects are classified and how educational poverty is distributed within the sample.

The analysis was based on an adaptation of the AMPI methodology (Adjusted Mazziotta-Pareto Index), already adopted by both Save the Children and ISTAT for the construction of indices at the regional and territorial levels (Save the Children, 2014; Pratesi, 2020; Mazziotta, 2024). AMPI is a non-compensatory method that aggregates standardized indicators through an arithmetic mean adjusted by an internal penalty function, assigning equal weight to each indicator. This methodology aligns with a vision of educational poverty as a multidimensional deprivation, where imbalances in even a single dimension can significantly impact the overall result.

In the present study, the AMPI methodology was adapted to an individual level of analysis: whereas the original indices were constructed at the regional scale (with reference to an "average" region), in our study the benchmark is the "average individual" within the sample. Accordingly, the score assigned to each subject reflects their level of educational deprivation relative to the project context.

To assess the consistency between the two indices, we compared the individual scores and conducted a linear regression to measure their correlation. The results were graphically represented using frequency distributions and scatterplots, in order to highlight overlaps and discrepancies in the classifications.

In addition to comparing the scores and classifications produced by the two indices, a predictive model was developed to explore the extent to which variables identified as relevant in the literature are associated with educational poverty. Two multiple linear regressions were estimated: the first included only ascriptive variables (gender, migrant background, parental education and employment status); the second also incorporated variables related to material and cultural resources (WEALTH, HEDRES, and the number of books in the home).

Results or Argumentation

The descriptive analysis revealed that the two indices yield different score distributions. Although a significant correlation was found between the scores obtained with the STC and ISTAT indices ($r = 0.62$),

the overlap between classifications is only partial. Some individuals identified as deprived by one index were not classified as such by the other, highlighting discrepancies in the identification of the phenomenon depending on the methodological choices adopted.

Graphical representations, including histograms of score distributions and scatterplots, illustrated these differences, showing how the relative position of individuals can vary considerably depending on the index used.
The proposed regressions based on the predictive model revealed that the educational poverty index is significantly associated—with both STC and ISTAT measures—with specific variables such as parental education level and the mother's employment status.
The predictive analysis shows that both indices are influenced by contextual variables, though with differing sensitivities: the STC index appears more responsive to the mother's education level and the type of school attended, while the ISTAT index shows a stronger association with material resources and the number of books at home. In both models, the availability of educational resources in the home (HEDRES) emerges as a consistent and significant predictor.
Overall, the results highlight the importance of carefully considering the definitional and methodological choices made in the construction of indices. These choices not only shape the statistical representation of the phenomenon but also affect the capacity for analysis and evaluation within educational interventions and the evidence base for compensatory policy design.

Keywords: Educational poverty, Multidimensional indices, Individual-level data, Comparative measurement

Preschool Attendance and Long-Term Learning Outcomes for Italian Children: An Exploratory Study Using INVALSI Data

Lorenzo Maraviglia – Cristina Stringher

Introduction
Preschool quality is increasingly recognized as a key factor for children's optimal cognitive, social, and emotional development, with consequences that can extend into later stages of life (EU Commission, 2014; Heckman, 2008; Melhuish et al, 2015; OECD, 2015). Research in this area underscores that early educational experiences can set developmental trajectories and have a significant bearing on learning and long-term academic performance.
In line with this evidence, the European Council has set a target of 96% attendance of quality preschool for eligible children by 2030 (EU Council, 2022). However, the question of what constitutes "quality" in early childhood education and care (ECEC) — and how it should be defined, measured, and monitored — remains highly debated in both the academic and policy literature (Eadie et al., 2024; EU Commission, 2018; McLean et al., 2022; Melhuish et al., 2015; OECD, 2015).
In Italy, preschool attendance rates are relatively high: in 2022 92,7% of 3–5-year-olds participated in preschool (EU Education and Training Monitor, 2024, still under the EU target of 96% for 2030). Yet, there is a relative lack of robust empirical evidence linking preschool experiences to measurable educational outcomes later in school. This study addresses this gap by exploring whether existing data, particularly from INVALSI assessments, can be used to assess the long-term impact of preschool attendance on pupils' learning outcomes in primary school.
Research Question and Theoretical Framework
Our study is conceptually anchored in a recent framework proposed by Stringher (submitted), which identifies at least eleven domains of ECEC quality, including dimensions such as quality for children, parents, staff, stakeholders, system-level quality, and both structural and process setting quality. Particularly

relevant to our analysis are two domains: *quality for children* and *long-term quality*, the latter being defined as the enduring outcomes of early education as reflected in later academic performance.
Building on this framework, we adopt a child-centred perspective that emphasizes the ethical imperative of educational quality, starting from the principle *primum non nocere* — first, do no harm. Our core hypothesis is that, at a minimum, preschool experiences should not negatively affect children's long-term educational outcomes. Positively, we are interested in whether preschool attendance is associated with improved academic performance during primary school, controlling for background characteristics.
Therefore, the central research question is to what extent does preschool attendance influence pupils' learning outcomes in grades 2 and 5 of primary education in Italy, as measured with standardized student assessments. In doing so, we exploit a decade of educational measurements carried out by INVALSI in these grades for national language and mathematics, so to comprehend trends in preschool impact.

Data

To investigate this research question, we utilize ten years of INVALSI individual-level data on pupils in the second and fifth grade for subsequent cohorts. These datasets include key variables at census and sample level: standardized achievement scores in Italian and mathematics, a binary indicator of preschool attendance, and contextual data on family background such as parental education, occupation, and citizenship status.
We focus on primary school outcomes because potential long-term effects of preschool are more likely to be detected before additional educational and contextual variables begin to exert stronger influences. Consequently, INVALSI data at the primary level allows us to capture early educational achievements that may be more directly attributable to preschool experiences.
It is important to note that information about preschool attendance is not consistently available across the datasets: on average, 20–25% of records lack this information, with variations by year and cohort. This presents a potential source of bias that must be addressed through appropriate statistical techniques.

Method

Our methodological approach involves multiple stages. First, we assess the implications of missing data on preschool attendance using various statistical techniques. Specifically, we apply Heckman selection models (Heckman, 1976) and Bayesian multiple imputation strategies (McElreath, 2020) to estimate whether the absence of data is likely to introduce significant selection bias. Preliminary results indicate that while some non-random patterns exist, the bias introduced appears limited.
Once data quality concerns are addressed, we proceed with multilevel regression modelling. These models estimate the effect of preschool attendance on INVALSI standardized scores in Italian and mathematics, while controlling for a rich set of covariates. These include gender, parental education level, socio-economic status, geographical location, and migrant background. This multilevel framework allows us to account for both pupil-level and school-level variability.
Crucially, we also explore effect heterogeneity by analysing whether the relationship between preschool attendance and academic outcomes differs across subgroups — particularly by region, socio-economic status, and immigrant background. Finally, given the observational nature of our data, we conduct sensitivity analyses to assess the robustness of our findings to unobserved confounding factors/variables.

Preliminary Results

Our preliminary analysis of missing data patterns suggests that, while not trivial, the lack of preschool attendance data is unlikely to invalidate overall findings regarding its impact on academic outcomes.
Regression analyses reveal that the average effect of preschool attendance on INVALSI achievement scores is modest but statistically significant at second-grade level. However, this effect tends to diminish by fifth grade, suggesting that early educational advantages may attenuate over time as other influences accumulate.
However, disaggregating by sub-populations, we find still relevant potential impact on achievements of qualified groups such as first and second generation foreign students. This finding suggests that preschool may play a compensatory role for vulnerable student populations, potentially contributing to equity in educational outcomes.
We also observe some discrepancies between results obtained from INVALSI sample data versus population data, which merit further methodological investigation. For example, we observe J curves for data collected after the COVID-19 pandemic (2021, 2022, 2023).

Relevance for INVALSI

Our study has direct implications for INVALSI and its role in educational monitoring and policy evaluation. First, it highlights the potential of existing INVALSI datasets to inform research on early childhood education

and long-term learning outcomes. However, it also reveals certain limitations in data completeness—especially regarding preschool attendance—which, if addressed, could significantly enhance the utility of INVALSI data, also for longitudinal research.
In this regard, we provide considerations and advice to improve data quality pointing to some enhancements that could further increase the value of this statistical source for social and educational research. First, dosage of preschool in terms of years attended by children is a key variable in international literature which could be collected from schools' administrative sources. Second, when process quality data from census preschool self-evaluations become nationally available, thanks to a collaboration between INVALSI and the Ministry of Education, data on children's outcomes in primary education could be regressed against data on preschool quality. This operation should allow us to precisely identify key factors contributing to both long-term and present preschool quality at the national level. Such a pioneering statistical infrastructure stretches educational research potential to inform evidence-based policy making.

Keywords: Preschool, long-term quality, INVALSI student assessments, trend analysis

Measuring and Mapping Educational Poverty: survey results among youth in Naples and the province

Rosaria Romano – Rosa Fabbricatore – Cristina Davino

Educational poverty (EP) is one of the most severe forms of poverty due to its long-lasting impact on individuals. It compromises their present and future, especially for adolescents, leading to negative consequences for society (Pratesi et al., 2021; Glaesser, 2022). Moreover, EP perpetuates a vicious cycle of disadvantage, fuelling the production of inequalities and their transmission across generations. Despite its fundamental relevance, measuring EP remains an open challenge in the scientific community, given its complex and multidimensional nature.
The goal of the contribution is twofold. Firstly, it aims to present a conceptual model that encompasses the primary factors contributing to measuring educational poverty in terms of deprivation or lack of opportunities, as well as the key outcomes of such deprivation at the individual level. The proposed operationalisation of EP is based on the definition provided by the NGO Save the Children (2014): "EP is the deprivation of children and adolescents from the opportunity to learn, experience, develop, and freely cultivate their capabilities, talents, and aspirations". Secondly, the contribution aims to discuss the main results of a survey conducted by the University of Naples Federico II in collaboration with Save the Children on youth aged 15-19 living in the city of Naples and its province. The empirical research focuses on young people aged 15-19 because they are at a critical development stage, where EO have long-lasting impacts on their future cognitive and social-emotional development (Heckman et al., 2013). The survey was conducted on a random sample of 3,200 students attending high school, stratified by area of residence, type of school, and school year.
The proposed conceptual model is based on the following research hypotheses:

(1) Students' social backgrounds significantly influence educational deprivation.

(2) EP, marked by limited access to educational opportunities (EO) in family, school, and environmental contexts, impacts cognitive (e.g., school performance) and non-cognitive abilities (e.g., self-esteem, motivation).
(3) A further outcome of EP is the capacity to aspire (Appadurai 2004), a culturally shaped individual's ability to envision and navigate pathways toward a better future.
(4) EO, EP outcomes and the interconnections among them are strongly affected by the place of residence, indicating the territorial heterogeneity of the phenomenon.

The research results make a significant contribution to the debate on measuring educational poverty. This detailed study measures the phenomenon at an individual level and considers numerous dimensions of the multidimensional concept, along with various outcomes. The proposed approach aligns with the framework proposed in 2024 by the ISTAT commission on measuring educational poverty However, two crucial differences should be highlighted: 1) ISTAT's reference population includes children and young people aged 0-19, while our study focuses specifically on students aged 15-19; 2) The indicators proposed by ISTAT are derived from official statistics and thus have a territorial reference, whereas our indicators are designed to assess EP at an individual level, providing a more granular understanding of the phenomenon. However, the study can be seen as an initial empirical result that could be extended to other territories and age groups.

Keywords: Educational opportunities, school performance, non-cognitive abilities, high school

Educational Accessibility and University Mobility: Evidence from the Italian Context

Cristian Usala – Iacopo Moreschini – Mariano Porcu – Isabella Sulis

This study investigates the relationship between local educational supply and student mobility choices in the Italian higher education system, addressing critical questions about territorial marginalization and demographic dynamics. Using data from the MOBYSU.IT database covering the population of students enrolled in Italian universities between 2019-2022, combined with georeferenced information on 53,868 schools nationwide, we examine how educational accessibility influences students' decisions to pursue tertiary education outside their home territories. We propose a composite indicator (I_m) measuring educational supply and accessibility for each municipality, incorporating both local high school availability and travel distances to nearest institutions. Students' mobility choices are classified according to the distance travelled to reach the chosen university and the availability of institutions in their local areas. Using logistic regression analysis, we find a significant negative relationship between local educational supply and mobility choices. Students from inner areas, with limited educational infrastructure, are more likely to choose distant institutions, suggesting that relative distances and mobility costs weigh less in their decision process, as attending tertiary education already implies significant travel. This effect is particularly pronounced in southern Italy. The findings highlight how territorial disparities in educational supply affect educational choices, contributing to demographic imbalances and potentially accelerating depopulation in marginalized areas. The results have important implications for educational policy, particularly regarding the equitable distribution of educational resources across territories and the challenge of maintaining viable communities in rural and inner areas.

Keywords: University mobility, territorial disparities, administrative data, inner areas

Educational poverty in the Italian Social Districs

Miria Savioli – Elisabetta Segre

This work is developed within the conceptual framework for measuring educational poverty developed by a scientific commission established in 2023 at Istat. The Commission's work was guided by a multidimensional approach that focuses on the broader meaning of the word 'education' in Italian when compared to the English term, broadening the view to include a variety of aspects that go beyond the failure to achieve educational goals. The framework is structured around two domains: resources and outcomes. The first domain concerns the educational/cultural resources available in the living context (family, school, places of learning and aggregation, etc.) and the opportunities that these resources offer in terms of experiences useful for personal growth. The second domain concerns the acquisition of cognitive and non-cognitive skills (emotional, relational, trust-based interactions) that enable individuals to grow and develop relationships with others, cultivate their talents, and actively and consciously exercise their right to citizenship. The domains are in turn structured into dimensions and sub-dimensions. Based on this conceptual framework, a set of multi-source indicators useful for measurement is currently being identified. During the Commission's work, some preliminary measurement exercises were carried out to identify the main critical issues regarding the use of available data sources, the territorial level, and the aggregation method. In particular, in 2024, some experiments were carried out to measure the phenomenon synthetically, through the calculation of two synthetic indicators, one for the domain of outcomes and one for that of resources at a territorial level. The composite indexes have been computed starting from a set of basic indicators available at municipality level. The territorial level of analysis has been, in one case (Istat 2024a), the intersection between the region and the degree of urbanization of municipalities and in two other cases (2024b, Pratesi et al 2025) the metropolitan city and some of its internal divisions. In all exercises, the AMPI method is used to construct the composite indices. In order to calculate the selected indicators at the territorial levels described, it was necessary to start from sources that provided data at the municipal level, which is why it was necessary to exclude all indicators that could be obtained from sample surveys. This led to the identification of a limited set of indicators calculated from the Permanent Population Census (ISTAT) and the MIM and INVALSI archives.

On the resource side, in order to assess the family context, indicators of the parents' socio-professional and socio-cultural status (including employment status and level of education) were considered, while in order to assess the resources provided by the school system, the adequacy of educational services and the school environment (overcrowded classrooms, architectural barriers for disabled students, lack of facilities such as computer rooms, gyms, auditoriums, canteens) and the actual use of educational services (including the number of children enrolled in nursery/preschool, use of the canteen, full-time attendance). Finally, to assess the social and cultural context, we used indicators relating to the living environment (degradation, urban green spaces, presence of libraries, museums, etc.), the presence of cultural facilities such as libraries and theaters in the municipality.

In terms of outcomes, cognitive skills are well monitored by INVALSI performance tests, which are administered to all students enrolled in the second, fifth, eighth, tenth, and thirteenth grades. Based on this source, indicators on literacy and numeracy levels were also developed, along with indicators on school dropout and school discontinuity based on Ministerial data.

Although the set of indicators obtained is not sufficient to provide an exhaustive quantitative representation of the conceptual framework prepared by the Commission, it has been useful in exploring the main methodological challenges to be addressed in order to arrive at a summary measurement of the phenomenon.

One of the aspects that emerged from these exercises regards the optimal territorial level. In particular, the inadequacy of the municipal level as the minimum territorial unit of analysis became apparent, so much so that the exercises, although starting from municipal databases, used aggregations of municipalities as the level of analysis. As is well known, this administrative unit places territories that are extremely different in size and complexity on the same level. While for larger municipalities, territorial boundaries can easily be too extensive to assess the actual accessibility and usability of facilities and services, for smaller municipalities, boundaries can be too narrow. In fact, there are many very small municipalities without schools, libraries, or museums. This does not mean that children and young people cannot have access to these services with travel times that are much shorter than those recorded in large cities. It is no coincidence that for smaller municipalities, it is the legislator itself that requires municipalities to form associations in

so-called social territorial areas (ATS) for the provision of social services. It is precisely from this administrative geography, regularly monitored by ISTAT as part of its survey on social services, that the idea for the work proposed in this abstract was born. Starting from the same set of census/administrative data, a summary representation of the phenomenon at the ATS level (Social Districts) is obtained. The ATS are approximately 600 aggregations of municipalities in which the planning and provision of social services and certain educational services are organized, so they can be representative in describing the phenomenon of educational poverty.
As regards the aggregation method, the use of the Alkire-Foster M0 index (Alkire et al. 2015) applied to individual territorial units is explored. The Alkire-Foster method involves a dual-cutoff mechanism: a first threshold identifies deprived units at the level of a single indicator and a second threshold identifies units in multidimensional poverty. Although the method is normally applied to individuals, there are no contraindications to its use on statistical units consisting of territories. It is one of the most widely used methods for the synthetic assessment of multidimensional deprivation and is characterized by several desirable properties, in particular decomposability.

The expected results are manifold. First of all, the aim is to test the adequacy of Social Districts in producing interesting readings of the phenomenon and their consistency with those that emerged from previous exercises. Furthermore, the application of the Alkire-Foster method will make it possible to assess the weight of individual dimensions and certain population subgroups in overall levels of educational poverty.

Keywords: Educational poverty, territory, educational resources, educational outcomes

SESSION 7. INEQUALITIES IN EDUCATIONAL ATTAINMENT AND THE ROLE OF TEACHERS AND SCHOOL PRINCIPALS

ORGANIZER: UNIVERSITY OF PIEMONTE ORIENTALE - UNIVERSITY OF MILANO BICOCCA - UNIVERSITY OF BERGAMO

COORDINATOR: GIORGIA CASALONE - ALESSANDRA MICHELANGELI - JURGENA MYFTIU

21ST NOVEMBER: 4.15 P.M. - 6.15 P.M. {AULA 5 – RESEARCH 17}

Mathematics score determinants at Grade 5: missing values recovery and variable selection via knockoffs

Silvia Bacci – Emanuela Dreassi – Leonardo Grilli – Carla Rampichini

In Italy, the National Institute for the Evaluation of the Education and Training System (INVALSI) administers standardised tests to assess ability in Italian, Mathematics, and English to students of different grades. For a sample of students and teachers, INVALSI also collects information about students' characteristics and school characteristics. We aim to detect which variables play a relevant role in explaining pupils' Mathematics ability at the end of the 5th grade. We consider data collected in 2022-2023. Globally, more than 40 covariates at the individual level and more than 30 covariates at the school level are available. Moreover, these variables are of mixed type (continuous and categorical) and are affected by missing values. This complex data setting requires a specific strategy to tackle the model specification and fit issue.

Variable selection via Knockoff in the presence of missing values

The selection of covariates having a relevant effect on a response is a fundamental issue in assessing a statistical model. It is particularly challenging when numerous covariates are available. Indeed, different selection strategies may lead to different results with the risk of including in the model variables with null effects or, on the opposite, excluding variables with a non-null effect.

In the literature, several approaches exist for variable selection. Barber and Candès (2015) introduced the knockoffs approach, having the advantage of controlling for the False Discovery Rate (FDR), that is, the proportion of variables wrongly declared non-null, while maintaining a high power (PWR). In their original formulation and following developments (see, among others, Ren, Wei, and Candès, 2023), knockoff methods typically assume continuous covariates and complete data sets. However, the INVALSI dataset has various binary, ordinal and count student variables, with a percentage of missing values that can exceed 25%. Thus, standard knockoff methods are not suitable for application to INVALSI data. To overcome the limits of standard knockoff-based approaches, we propose a flexible strategy to treat missing data under the assumption of missing at random (MAR; Little and Rubin, 2002). In particular, we suggest to use a multiple imputation (van Buuren, 2018) technique (e.g., Multiple Imputation by Chained Equations - MICE) to obtain a complete dataset; moreover, to control for the variability due to the random nature of the multiple imputation techniques, we suggest repeating the imputation procedure a certain number of times (e.g., 10). Then, we perform the variable selection via the knockoff filter on each imputed dataset. In particular, we apply the sequential knockoffs approach of Kormaksson et al. (2021), recently optimised from a computational perspective by Zimmermann et al. (2024), that allows for the selection of mixed-type (i.e., continuous, binary, categorical) variables. Variables in at least 80% of cases are then selected and retained in the model fitting. The properties of our proposal have been investigated using a Monte Carlo simulation study (Bacci et al., 2025), confirming the good performance of our strategy in terms of FDR and PWR.

Determinants of Grade 5 mathematics test scores

We focus on data from the INVALSI sample survey of the year 2022-23 on pupils in Grade 5, corresponding to the last year of primary school. The dataset includes 16,828 students nested in 501 schools.

Specifically, we consider the achievement in mathematics, a latent ability measured by a set of dichotomously scored items. INVALSI provides the response pattern to the items and a Rasch-based estimate of the ability (Rasch, 1960). In this contribution, we use the ability estimate: in the sample, the average value is 191.82 points with a standard deviation of 41.07.

The data set includes the geographical area of the school and 30 student-level variables of different types:

- binary (e.g., gender, availability of a personal computer);
- ordinal (e.g., mother's education);

- unordered categorical (e.g., mother's occupation).

To study the relationship between the background characteristics and pupils' achievement, while accounting for their clustering into schools, we specify a random intercept model (Snijders & Bosker, 2011). Denoting the pupils with the index i (level 1) and the schools with the index j (level 2), the model is specified as:

$$y_{ij} = x_{ij}\beta + z_j\gamma + u_j + \varepsilon_{ij}, \qquad (1)$$

where y_{ij} is the mathematics score of pupil i of school j, x_{ij} is the row vector of level 1 predictors with coefficients β (including the intercept) and z_j is the row vector of level 2 predictors with coefficients γ. The level 1 and level 2 errors are assumed to be independent and identically distributed with a normal distribution, that is, $\varepsilon_{ij} \sim N(0, \sigma_\varepsilon^2)$ and $u_j \sim N(0, \sigma_u^2)$, respectively. Moreover, they are also assumed reciprocally uncorrelated.

The background variables are highly correlated, and there are no theoretical guidelines to choose among them, calling for a data-driven selection procedure.

Moreover, a critical issue of the data is the relevant rate of missing values for many student-level variables, ranging from 1.6% to 25.5%. Specifically, the variables with the largest rates of missing values are Father's and Mother's education and occupation (from 23.4% to 25.5%), followed by Italian language at home, Dialect spoken at home, Italian used with friends, and Nursery school (about 10%). A complete-case analysis would halve the number of students with a large loss of power and, especially, a potentially high bias given that the missing mechanism cannot be assumed to be missing completely at random. Rather, a MAR mechanism with missingness depending on observed values may be plausible, so multiple imputation is a viable approach to preserve the original sample size and avoid bias in the estimators.

To perform variable selection while handling the missing data issue, we follow our proposed procedure described in Section 1. Specifically, we implement MICE with 10 imputed datasets using the mice package of R (van Buuren and Groothuis-Oudshoorn, 2011). Each variable is imputed with a model suitable for its measurement scale: logit model for binary variables, cumulative logit model for ordinal variables and multinomial logit model for unordered categorical variables.

For each imputed dataset, the predictors are selected using the sequential knockoffs procedure of Kormaksson et al. (2021), as implemented in the R package knockofftools (Zimmermann et al., 2024) with options PFER=2 and method="sparseseq". Unordered categorical and ordinal variables are coded using a set of dummy variables with a baseline category. In the knockoffs filter, we use group lasso (Yuan and Lin, 2006) instead of standard lasso, to avoid the selection depending on the arbitrary choice of the baseline category. In this way, the corresponding set of dummy variables is jointly selected or discarded for each predictor. The procedure runs 31 times for each imputed dataset, and a predictor is retained if selected in at least 50% of the runs.

Table 1 shows the predictors with the proportion selected over the 10 imputed datasets by the sequential knockoffs procedure. According to the findings of a simulation study (Bacci et al., 2025), a predictor is included in the final model if selected in at least 8 out of 10 imputed datasets: this is denoted by a check-mark in Table 1. The School area and 22 out of 30 student-level variables have been selected.

The sequential knockoff procedure gives parameter estimates based on the group lasso for a standard linear model, i.e. without random effects. This is not a limitation for selecting the predictors since the knockoff statistic is based on the estimated regression coefficients, which are minimally affected by the random effects since, in the linear model, conditional and marginal coefficients coincide. Notwithstanding, we need to fit a random effect model with the selected variables for two reasons: (i) to obtain standard errors properly accounting for the school clustering; (ii) to avoid the downward bias implicit in lasso-based procedures (e.g., Belloni and Chernozhukov, 2013). For these reasons, we fit the random intercept model (1) with the selected predictors to each imputed dataset via maximum likelihood using the R package lme4 (Bates et al., 2015). The results across the 10 imputed datasets are combined using Rubin's rules to obtain point estimates and standard errors.

Table 2 reports the residual variances and the Intraclass Correlation Coefficient (ICC) for the model without predictors (null model) and for the model with the predictors selected by sequential knockoffs (full model; see Table 1). The null model shows that the variability in the mathematics scores due to school clustering is 13.8%. The predictors explain about 23% of the school-level variance and about 15% of the pupil-level variance. Given the included predictors, the residual ICC in the final model is 12.6%, indicating that the school clustering is still relevant.

Table 3 presents the point estimates and corresponding standard errors for the full model. The effects align with expectations, with a few notable exceptions. Interestingly, personal tablet or smartphone ownership is associated with lower mathematics scores. Since the test evaluates mathematics proficiency acquired during primary school, this result could reflect the potentially harmful impact of device usage by children (Beland & Murphy, 2016). Additionally, owning such devices may signal a family environment less conducive to learning.

Mothers' and Father's education have similar effects, monotonically increasing from compulsory to master or more. On the other hand, the effects of occupation are much greater for the father. The largest effect pertains to the number of books at home, which is a proxy of the cultural environment.

Final remarks

In this contribution, we analysed the determinants of math scores at grade 5, proposing an approach for selecting mixed-type variables in the presence of missing values using knockoffs. Our proposal presents two main advantages: (i) the use of multiple imputation to deal with missing values, allowing to deal with any type of variables under the MAR assumption, and (ii) the selection of variables based on the sequential knockoffs procedure of Kormaksson et al. (2021) and Zimmermann et al. (2024) that allows to generate knockoff copies of unordered categorical variables based on the multinomial logit model, differently from standard knockoffs procedures, which are designed for continuous variables.

The results confirm the effectiveness of the sequential knockoffs procedure in identifying key predictors of mathematics achievement, with 22 student-level variables and the school area selected. Fitting a random intercept model allowed us to account for school-level clustering and correct the bias inherent in lasso-based methods. The selected predictors explain a substantial portion of the variance at both the school level (23%) and the student level (15%), although residual clustering effect remains relevant (residual ICC of 12.6%). The estimated effects are generally consistent with expectations, highlighting the importance of the family and cultural environment (e.g., number of books at home). At the same time, the negative association between device ownership and math performance raises important questions about the educational impact of early technology use.

A general issue about knockoffs, unrelated to missing data, is their performance in selecting cluster-level predictors in multilevel analysis. The available methods to generate the knockoffs do not account for the levels of the predictors, so a copy of a cluster-level variable is produced as if it were an individual-level variable, ignoring the constraint that it has to be constant within the clusters. This limitation is negligible in the application presented here because all predictors, except the geographical area, are at the student level. Future research should develop a strategy for the selection of cluster-level predictors to consider teacher and school characteristics available in the INVALSI sample data.

Table 1 Sequential knockoff variable selection: proportion selected over 10 imputed datasets and retained status (✓ if proportion ≥ 0.8). INVALSI Grade 5 sample survey, year 2022-23.

Variable	*Levels*	*Pr.*	*Ret.*
Male	1=yes, 0=no	1	✓
# brothers	ordinal, 5 levels	1	✓
# sisters	ordinal, 5 levels	1	✓
At home, you have			
quiet space	1=yes, 0=no	0.9	✓
computer	1=yes, 0=no	1	✓
personal desk	1=yes, 0=no	0.3	
software	1=yes, 0=no	0	
internet connection	1=yes, 0=no	1	✓
personal room	1=yes, 0=no	1	✓
classical books	1=yes, 0=no	0	
artworks	1=yes, 0=no	0	
technical manuals	1=yes, 0=no	0	

dictionary	1=yes, 0=no	1	✓
personal tablet	1=yes, 0=no	1	✓
smartphone	1=yes, 0=no	1	✓
# books	ordinal, 5 levels	1	✓
Student origin and language			
Student born in Italy	1=yes, 0=no	1	✓
Immigrant staus	1=native; 2=1st gen immigrant; 3=2nd gen imm.	0	
Italian language at home	1=yes, 0=no	1	✓
Dialect spoken at home	1=yes, 0=no	1	✓
Italian used with friends	1=yes, 0=no	1	✓
Parents			
Father born in Italy	1=yes, 0=no	0.2	
Mother born in Italy	1=yes, 0=no	1	✓
Father's education	1=compulsory, 2=high sch., 3=college, 4=master+	1	✓
Mother's education	1=compulsory, 2=high sch., 3=college, 4=master+	1	✓
Father's occupation	1=unemployed 2=manager, clerk 3=self-employed, 4=blue collar 5=professional 6=retired	1	✓
Mother's occupation	1=unemployed 2=manager, clerk 3=self-employed, 4=blue collar 5=professional 6=housewife, retired	1	✓
Student career			
Nursery school	(for pupils aged 3-5 years) 1=yes, 0=no	1	✓
Regular student	1=yes, 0=no	1	✓
School weekly hours>30	1=yes, 0=no	0	
School area	1=NW; 2=NE; 3=Center; 4= South; 5= S-Islands	1	✓

Table 2 Random intercept null model and full model with selected predictors (Table 1): combined estimates of the residual variances based on 10 imputed datasets. INVALSI grade 5th sample survey, year 2022-23.

Residual variances	null model	full model	% explained
Level 2 (school)	233.30	178.94	23.30
Level 1 (pupil)	1,461.94	1,239.97	15.18
Total	1,695.24	1,418.91	16.30
ICC	13.76%	12.61%	

Keywords: Equality of opportunity, education returns, earnings, policy evaluation

Equity or Efficiency? The Role of School Principals in Mediating Accountability Pressures

Francesco Balzaretti – Tommaso Agasisti – Moris Triventi

In recent years, the growing emphasis on administrative accountability systems has led many countries to implement mechanisms aimed at enhancing responsibility within the education sector. The primary rationale underlying these policies lies in the pursuit of greater efficiency, effectiveness, equity, and good governance (Verger & Parcerisa, 2017).

However, these accountability pressures on schools are not without cost. Scholars have argued that accountability policies may restrict teachers' professional autonomy, shifting their roles toward more administrative and technical functions (Day, 2002), and unintended consequences on pedagogy, learning process and teacher-student relationships have been identified (Jones & Egley, 2004; Valli & Buese, 2007).

Moreover, school principals play an important role in mediating the impact of accountability reforms at the school level, with their leadership styles shaping how such policies are interpreted, enacted, and experienced by teachers (Diamond, 2007). Far from being passive conduits, principals exercise agency: as Finnigan (2012) emphasizes, their responses to accountability mandates can mobilize teachers toward performance-oriented practices.

The present research seeks to contribute to this discourse through an original analysis based on a survey conducted by INVALSI with a sample of Italian school principals.

The questionnaire, completed by principals from a representative sample of schools participating in the national standardized assessment, contains information that can be grouped into five clusters: principals' attitudes toward standardized testing, stakeholder engagement in school matters, contextual characteristics of the school, personal background of the principal, and school management practices (Agasisti, 2020).

The research investigates how principals' attitudes toward standardized assessments, a dimension encompassing both perceptions and use of assessment results, shape school performance, particularly in relation to the socioeconomic composition of the student body. The underlying premise is that a principal's strong endorsement of standardized testing may not only increase overall achievement levels but also interact with structural inequalities within the school. For instance, such leadership stances could inadvertently exacerbate achievement gaps among students by channeling resources on overperformer students, thereby deprioritizing those with greater learning needs.

This study thus aims to unravel the mechanisms through which principals' orientations towards accountability influence both performance and equity, deepening our understanding of how Italian school principals translate the accountability mandate into educational practices, in light of preliminary evidence suggesting that they actively engage in negotiating external demands (Paletta et al., 2019; Paletta, 2019; Mentini, 2024). Moreover, research indicates that, in the Italian context, principals' managerial practices have a greater impact on disadvantaged groups of students, suggesting that principal effectiveness plays a major role where it is most needed (Agasisti, 2020).

Finally, we argue that a close examination of the mechanisms through which mandated accountability programs influence broader educational processes has relevant implications for the design of effective policies that aim to reconcile the pursuit of efficiency with the goal of fostering more equitable education systems.

Keywords: School Principals, Accountability, School performance, Inequalities.

The role of high schools' performance on inequalities in educational attainment and student mobility

Maria Berrittella – Francesco Salomone Marino

Introduction

Previous studies have shown that students graduating from scientific and classical lyceum are expected to be more likely enrolled at northern universities, as well as there is a positive relationship between student mobility and individual skills, which leads to the so-called brain drain from the South (Faggian et al., 2017; Tosi et al., 2019; Attanasio et al., 2020; Rizzi et al., 2021; Usala et al., 2023). Selectivity emerges in student mobility, which is related to the family background: students coming from a wealthy and stimulating family environment with greater financial parental support and high cultural resources of families are more likely to move to northern regions for tertiary education (Grodsky et al., 2010; Ballarino et al., 2016; Impicciatore et al., 2019; Usala et al., 2023). Parental education investment on student mobility is perceived as an opportunity for reinforcing the families' social status (Ballarino et al., 2016; Tosi et al., 2019). In this rich literature on student mobility, the role of the high school has been related to the type of high school track of the student. The common result is that the South to North movers are mainly those students coming from the lyceum rather than the students coming from technical or vocational schools. However, one aspect that has been neglected in literature on student mobility is the schooling effects, that do not depend on the type of high school track, but they should be related to a measure of the performance of the high schools. In this context, this study explores the existence of inequality in the students' skills from Invalsi test scores and its determinants for Southern students; then this study investigates how these inequalities in secondary schools can affect the student's choice for tertiary education. In particular, we focus on student mobility from the South to the northern universities. Moreover, we investigate if the performance of the secondary education schools of origin can weaken or strengthen the inequalities and if it can affect student mobility.

Aims and research hypothesis

The main aims of this paper are two. First, to investigate the relationship between secondary schools of origin, inequalities in students' skills and student mobility. Second, under which conditions the high schooling effect prevails on the socio-economic status of the family.

Data

The data used in this paper are drawn from the Database MOBYSU.IT (2020), which has been processed in accordance with the research protocol for the study "From high school to the job placement: analysis of university careers and university mobility from Southern to Northern Italy" among the Ministry of University and Research, the Ministry of Education and Merit, the University of Palermo as the lead institution, and the INVALSI Institute. It merges information from three national administrative archives: (i) the National Archive of Schools (ANS-S), which contains micro-data of Italian high school students from 2014-15 to 2018-19; (ii) the database of the National Evaluation Institute for the School Systems (INV), which contains micro-data of the students participating to the INVALSI tests, that are standardized national assessment tests to evaluate the students' performance, at different levels of education, in Mathematics, Italian language, English reading and listening; (iii) the National Archive of University Students (ANS-U), which contains micro-data comprehensive of all the information about the university careers of all the students enrolled in Italy from 2008 to 2020.

Methods

We apply multinomial logistic regression models to longitudinal data on high school university transition regarding southern students enrolled and not enrolled at university in Italy. We define the dependent variable by five levels: not enrolled at university (not enrolled); enrolled at university in the region of residence (stayer), enrolled at university in a different southern region (South to South mover), enrolled in a region in the Centre of Italy (South to Centre mover) and enrolled in a region in the North of Italy (South to North mover). The set of control variables includes secondary school track, Invalsi test score, high school grade, region of residence and fields of study. Therefore, we consider the family background in terms of socio-economic status and the performance indicators related to the secondary school of origin. We construct high school performance indicators to explore the high schooling effect on students' inequalities and if it is higher than the effect of the family background.

Results

We estimate the odds ratios, which measure the strength of the relationship between the predictor variables and the probability of being a mover. We control self-selection, related to the competencies and choices of

the southern students, such as skills, high school track and fields of study, and other selection mechanisms, such as gender, geographical origins and family background. Then we investigate the odds ratios related to the contribution of the performance of high schools to inequalities in Invalsi test score and student mobility. Moreover, we compute the average predicted choice probabilities and, subsequently, the average marginal effects. The average predicted choice probability is estimated using the coefficients obtained from the multinomial logistic regressions associated with the student's status (not enrolled at university, stayer, South to South mover, South to Centre mover, South to North mover). We separately get predicted probabilities for male and female students. Moreover, following Long et al. (2021), we calculate the average marginal effects as discrete changes, that is, we compute group differences in average predicted probabilities, for example between male and female students, or for different parental socio-economic status. As in previous studies, our findings confirm that the South to North movers are mainly the most talented high school students, coming from the lyceum and from the middle-upper income families. Differently, we also find that a schooling effect exists: the students are more likely to be South to North mover if their secondary schools of origin have a high value in the school performance indicators. The schooling effect is stronger for the students with lower skills and coming from disadvantaged families.

Keywords: Secondary education, Performance indicators of the schools, Student university mobility, Socio-economic status

Why Do Some Schools Perform Better Than Others? Exploring the Black Box of the School Effect

Giorgia Casalone – Alessandra Michelangeli – Jurgena Myftiu

This study investigates the determinants of the school value added by analysing INVALSI test data from Grades 2, 5, and 8 in Italy. Using multilevel regression models, it examines how student achievement is influenced not only by individual background, but also by class composition, teacher characteristics, and principal managerial practices. Factor analysis is used to derive latent indicators of principals' leadership style and organizational approach. By decomposing the total variance into achievements into different levels, we assess the relative weight of individual, classroom and school influences on student learning outcomes. The analysis highlights the underexplored role of teachers and school principals in shaping learning outcomes.

1. Introduction

This study investigates the factors behind the "school effect" or "school value added", defined as the difference between students' actual performance and their predicted performance based on background characteristics. Analysing a nationally representative sample of Italian students who took standardized academic achievement tests in grades 2, 5, and 8 (INVALSI), the paper explores the impact of factors beyond individual characteristics, including the quality of classmates and the attributes of teachers and principals. Regarding teachers, the study examines gender, educational qualifications, experience, and contract type (fixed-term or permanent). For school principals, it considers leadership style (autocratic or democratic), organizational approach, and the school climate they help foster, assessing how these factors may affect students' learning achievements. The literature concerning the teacher value added is extensive (Carrel et al. 2010, Chetty et al. 2014, Barrios- Fernández and Riudavets-Barcons, 2024), whereas the contribution of school principals is less explored (Bartanen et al. 2024; Bloom et al. 2015), especially in Italy (Masci et al. 2017). Using multilevel regression models and factor analysis to uncover latent dimensions such as leadership style and school climate, the analysis investigates how these variables influence student outcomes and reduce the unexplained variance in learning achievements across classes and schools. The findings provide insights into the determinants of school effectiveness, thereby informing educational policy and school management strategies.

2. Empirical strategy

Multilevel analysis

We adopt a three-level hierarchical linear model to analyse student achievement in the INVALSI standardized tests in Italian and Mathematics across Grades 2, 5, and 8. The multilevel structure of the data,

where students (Level 1) are nested within classes (Level 2), and classes are nested within schools (Level 3), requires an approach that accounts for dependence across units at different levels. The outcome variable is the standardized test score, modelled as a function of covariates at all three levels.
The model specification is as follows:

$$Yijk = \beta0 + \beta1*Xijk + \beta2*Wjk + \beta3*Zk + u0k + v0jk + eijk \quad (1)$$

Where Yijk denotes the test score of student i, in class j, in school k, Xijk is a vector of student-level variables (e.g., gender, nationality, socioeconomic status, prior achievement), Wjk refers to class-level characteristics (e.g., peer composition, timetable structure, teacher gender, age, years of experience, academic qualification, and contract type) Zk includes school-level covariates (e.g., geographical location, principal's gender, age, tenure in the school, leadership style: collaborative vs. autocratic, and activism). u0k and v0jk are the random effect at the school and class level, eijk is the student-level residual
This model decomposes the total variance in achievement into three components: between schools, between classes within schools, and between students within classes. The intraclass correlation coefficients derived from the estimated variance components quantify the proportion of variance attributable to each level. This allows for the assessment of the relative weight of individual, classroom, and school-level influences on student learning outcomes. By explicitly modelling contextual factors at multiple levels, this approach provides a robust framework for understanding the determinants of educational achievement and for informing targeted interventions at the appropriate organizational layer of the school system.

Factor Analysis
To capture key dimensions of school principals' leadership style and activisms, we apply factor analysis to a set of questionnaire items addressed to principals in the INVALSI surveys. These items cover a broad range of behaviours and attitudes, including decision-making approaches, delegation practices, staff coordination, goal setting, communication with teachers, and overall management philosophy.
Given the high dimensionality of the raw survey data and the theoretical expectation that underlying latent constructs (e.g., leadership style, organizational strategy) drive observable responses, factor analysis allows us to reduce complexity and derive synthetic indicators that summarize these aspects in a statistically rigorous way. The analysis is conducted separately for items intended to reflect managerial leadership style and those related to organizational techniques.
Once the factors were extracted and validated through standard procedures, including the Kaiser-Meyer-Olkin measure of sampling adequacy, Bartlett's test of sphericity, and the inspection of factor loadings, the resulting standardized factor scores were used as continuous covariates at the school level in the multilevel model.
These latent indicators enable us to assess the association between different styles of school leadership and organizational management, ranging from shared to more hierarchical or bureaucratic models, and student achievement outcomes across different educational stages.

3. Dataset description

We use INVALSI sample data from the school years 2017/2018 to 2022/2023, excluding 2019/2020, when testing was suspended due to the COVID-19 pandemic. We focus on Grades 2, 5, and 8 in order to examine whether and how the characteristics of teachers and school principals influence student outcomes across different educational cycles. The sampled data INVALSI are the only ones that provide detailed information on students, teachers, and school principals.
Teachers and principals in the sampled classes and schools were asked to complete a questionnaire, which varies slightly from year to year. The questionnaire collects both demographic information (e.g., gender, age, years of service) and details on academic qualifications, including degree type and field of study. In addition, depending on their role (teacher or principal), respondents answer a set of items aimed at capturing how they perform their professional duties. Of particular interest are the numerous items addressed to school principals, which make it possible to reconstruct their managerial and organizational leadership style. The number of principals by school grade and year is reported in Table 1.

Table 1: Number of principals by school grade and year

Grade	School year				
	2017/2018	2018/2019	2020/2021	2021/2022	2022/2023
2	552	487	449	428	416
5	552	489	449	430	419
8	541	545	223	86	74

Regarding gender (Table 2) the data highlight a divergence between school levels: while female representation in leadership roles continues to grow in primary education, it has declined significantlzy in secondary schools, particularly in the aftermath of the pandemic.

Table 2: Share of female principals by school grade and year

Grade	School year				
	2017/2018	2018/2019	2020/2021	2021/2022	2022/2023
2/5	.705	.732	.742	.768	.750
8	.682	.712	.691	.658	.570

Table 3 shows a generational shift in school leadership across both primary and lower secondary schools. The share of older principals (61–67) declined steadily after 2018/2019, while the 49–60 age group remained the most represented. Notably, the proportion of younger principals (18–48) increased in the recent years, especially in primary schools. These trends suggest a gradual leadership renewal, more pronounced at the primary level.

Table 3: Principals' class age by school grade and year

Grade	Class age	School year				
		2017/2018	2018/2019	2020/2021	2021/2022	2022/2023
2/5	18-48	19.63	11.37	20.22	19.73	29.58
	49-60	50.01	47.25	52.83	58.06	54.25
	61-67	30.36	41.37	26.95	22.21	16.16
8	18-48	15.90	15.43	20.91	26.63	22.33
	49-60	51.34	45.70	53.42	52.61	55.62
	61-67	32.77	38.87	25.68	20.77	22.05

Table 4: Principals' education

Grade	Class age	School year				
		2017/2018	2018/2019	2020/2021	2021/2022	2022/2023
2/5	University degree	77.27	70.65	62.25	65.29	59.54
	Postgraduate degree	22.72	29.35	37.75	34.71	40.46
8	University degree	76.23	69.39	61.31	62.11	65.6
	Postgraduate degree	23.77	30.61	38.69	37.89	34.4

University degree includes: pre-Bologna single cycle degree (Lauree vecchio ordinamento), bachelor degree (laurea triennale), master's degree (laurea magistrale), post-Bologna single cycle degree (Lauree a ciclo unico)

Postgraduate degree includes: 2nd level master degrees (master di secondo lovello), PhD, specialist training (scuole di specializzaione)

Table 4 shows a steady increase in the share of principals with postgraduate qualifications in both primary and lower secondary schools. In primary schools, the proportion rose from 22.7% in 2017/2018 to 40.5% in 2022/2023; in secondary schools, from 23.8% to 34.4%, with a slight dip in the final year. This trend

suggests a growing professionalization of school leadership, likely driven by generational turnover and greater emphasis on advanced training post-pandemic.

Table 5: Principals' degree field of study by school grade and year

Grade	Class age	School year				
		2017/2018	2018/2019	2020/2021	2021/2022	2022/2023
2/5	Humanities	37.00	43.02	42.26	45.32	43.69
	Psycho_Pedagogy	27.22	23.86	22.88	17.89	24.09
	STEM	12.83	13.34	10.87	12.58	9.65
	Law_Socio_Economics	10.55	7.27	10.54	11.51	8.92
	Others	12.41	12.50	13.45	12.70	13.65
8	Humanities	38.85	37.43	50.67	47.19	45.87
	Psycho_Pedagogy	26.93	23.56	23.14	16.12	14.46
	STEM	20.14	18.26	8.66	12.13	17.76
	Law_Socio_Economics	7.60	10.96	13.00	12.87	4.77
	Others	6.48	9.79	4.53	11.68	17.14

Table 5 shows that principals in both primary and lower secondary schools mostly hold degrees in Humanities, though their share has slightly declined in recent years. Psycho-pedagogical backgrounds have decreased significantly. STEM degrees remain limited in primary schools but show some recovery in lower secondary. Overall, there's a gradual shift toward more diverse academic profiles among school leaders.

Table 6: Years leading the current school (mean and S.D.)

Grade	School year				
	2017/2018	2018/2019	2020/2021	2021/2022	2022/2023
2/5	5.68 (5.71)	6.20(6.76)	5.76(6.33)	5.32(5.66)	5.64(5.98)
8	5.80(5.56)	5.30(4.48)	4.54 (4.86)	5.48(3.90)	4.30(4.38)

Table 6 shows the average years principals have led their current school. In primary schools (Grades 2/5), tenure remains relatively stable around 5–6 years. In lower secondary schools (Grade 8), a gradual decline is observed, especially in 2022/2023, suggesting more recent appointments or greater turnover. Standard deviations indicate high variability across both levels.

4. Principals' leadership and activism

Leadership: TBD

Regarding activism, the sampled principals were asked how frequently they engage in the following activities:

A1. I ensure that teachers' professional development activities align with teaching objective
A2. I ensure that teachers work in accordance with the school's educational objectives.
A3. I monitor classroom activities.
A4. I provide teachers with suggestions to improve their teaching.
A5. I oversee students' assessment activities.
A6. When a teacher faces difficulties in their class, I take the initiative to discuss the issue with them.
A7. I inform teachers about opportunities for subject-specific and teaching-related professional development.
A8. I make sure to work towards specific goals and/or according to the school's educational plan.
A9. I take final evaluations into account when making decisions about the current year's planning.
A10. I ensure that responsibilities for coordinating the school curriculum are clearly defined.
A11. I handle the resolution of disciplinary issues within the classes.
A12. I personally cover for teachers who are unexpectedly absent in class.

The possible responses are "Never or almost never", "Sometimes", "Often", "Always or almost always".

The factor analysis extracted one factor explaining 89% of the variance across 12 variables measuring principals' activity levels. This indicates that these variables are highly correlated and can be effectively

summarized by a single indicator of "activism." Factor loadings ranged from moderate to high (0.27–0.67), indicating that all items contribute positively to the construct of "activism," with some items more strongly related than others. Variables such as A2 (0.67), A8 (0.65), and A9 (0.65) are most strongly correlated with the factor, thus serving as better indicators of the principal's activism. A12 has the lowest loading (0.27), indicating a weaker contribution. The generated factor score (activism_factor) represents a synthetic continuous indicator of principal activism. This score can be used in multilevel models to assess how principals' activism influences student outcomes, simplifying the analysis by summarizing multiple related variables into a single meaningful measure.

Keywords: Learning achievement, Education production function, Education inequality, School effect

Towards Participatory Teaching: the State of Active Learning in Italy

Althea Bianca Dal Ben – Anna Loi – Ivan Blancato – Sofia Ercolanoni

In recent decades, the Italian school system has undergone several transformations. An increased awareness of the different learning mechanisms specific to each student has emerged, promoting not only new challenges but also opportunities for educational innovation.
At the same time, the changing conception of the student's role - from a passive recipient of information to an active participant in the construction of the learning process - has led to the introduction of teaching strategies that complement traditional lectures.
One of the main objectives behind the implementation of an alternative didactic approach is to enhance the learning that stems from experience, which results in a different and greater level of student engagement (Nigris et al., 2007). For this reason, alternative teaching methods are particularly effective for students with language barriers or special educational needs (D'Elia et al., 2025; Garrels, 2019).
There are several ways to integrate innovative elements into a traditional lesson. For this reason, Prince (2013) describes *active learning* as an educational approach rather than a single method.
The author argues that, although providing a single, definitive definition is complex, two core characteristics can be identified: the inclusion of activities carried out by students during traditional lessons and the promotion of their active engagement.
Three types of *active learning* are defined: *collaborative learning*, whose central element lies in the creation of student groups; *cooperative learning*, which is based on replacing competition among classmates with cooperation and mutual support; and *problem-based learning*, which emphasizes solving a given problem by encouraging student creativity and avoiding the imposition of predefined strategies (Prince, 2013).
The various forms of active learning, in their heterogeneity, stimulate the learner's direct action, the elaboration of encountered concepts, and self-reflection regarding any doubts. Although these activities are part of the traditionally conceived educational process, the innovative element lies in the fact that these processes are now stimulated during class time, rather than being postponed exclusively to an individual study phase at home. Thanks to active learning, these stages of the learning process are integrated into everyday teaching and become collective experiences within the classroom.
The present study aims to examine the diffusion of alternative teaching approaches within the Italian school context, with a twofold objective.
First, to analyze the characteristics of teachers who adopt these methods, as well as the distribution of active practices across different educational settings. The goal is to identify any specific traits of teachers that may be linked to the implementation of the methodologies under investigation, and to assess which students are most frequently exposed to them. The research hypothesis, in this regard, is that teaching seniority and precarious employment conditions may negatively affect the implementation of innovative practices. Likewise, the concentration of students at risk of educational disadvantage - whether in terms of socioeconomic background or migratory status - could result in more limited adoption.
Second, the study seeks to assess the effectiveness of these methodologies by investigating their impact on the learning outcomes of students exposed to them. The goal is to understand whether and to what extent such teaching methods can help close existing learning gaps. The research hypothesis is that the use of active methodologies is associated with improved outcomes, particularly for students who typically face greater challenges in traditional learning environments.

To achieve this goal, a dataset was constructed by integrating information collected from the national INVALSI surveys - specifically, from the teacher questionnaire - with student background data and learning outcomes. The adoption of this dataset made it possible to link teachers' responses with their students' standardized test results at grade 8. In particular, by combining data from different survey waves, a comprehensive database was created, allowing for in-depth analyses: the sample includes 600 classes, approximately 11500 students, and 1440 teachers.

The richness of this dataset allows for the investigation - within the broad category of active methodologies - of highly heterogeneous practices, such as: sharing educational materials through platforms/cloud services; activities involving discussion and conversation; peer-based activities (*cooperative learning*, *peer-to-peer*); assignments requiring students to present their work in various formats (slides, videos, multimedia projects, etc.); *flipped classroom* teaching; laboratory-based activities; and extracurricular remediation or enrichment activities. In this way, it becomes possible to distinguish the varying prevalence of each of the considered teaching practices, with the aim of first evaluating which ones are most widespread and, secondly, identifying which students are most frequently exposed to them, both in terms of individual characteristics (e.g., socio-economic background) and contextual factors (e.g., class size, proportion of students with a migrant background). These data also allow for the identification of the characteristics of the teachers who implement such practices, which are hypothesized to be related to the actual adoption of the methodologies under investigation.

The depth of the resulting database made it possible to investigate both of the previously mentioned research hypotheses. In the first phase of the analysis, the various distributions of active teaching methodologies were examined based on the information reported by teachers. This phase began by analyzing the distribution of these practices with respect to teacher characteristics, such as seniority within the school and contract type, as well as possible differences across subjects (Italian, mathematics, and English). This phase also considered contextual variability by looking at different geographic areas (North-West, North-East, Center, South and Islands) and the specific features of the classes, using an indicator of the average socio-economic and cultural background at both the school and classroom levels.

In the second part of the analysis, the association between the adoption of active methodologies and student performance in the INVALSI tests for Italian and Mathematics is estimated, controlling for the variability examined in the first phase. This association is explored by differentiating based on salient student characteristics, in order to assess the effectiveness of such methodologies for different student profiles.

Keywords: Active learning, Teaching practices, Learning outcomes, Teachers

The Gender Grading Gap in Italy: Heterogeneity, Trajectories, and Consequences from Large-Scale Population Student Data

Carlos J. Gil-Hernández – Giovanni Abbiati – Ilaria Lievore

Introduction

In recent decades, the educational landscape in OECD countries has undergone a significant transformation: girls, historically disadvantaged, now outperform boys in several educational indicators, including average grades and completion rates (DiPrete & Buchmann, 2013; OECD, 2024). However, gender disparities persist in STEM pathways, where boys tend to perform better in standardized tests while receiving lower teacher-assigned grades than girls in the same subjects (Voyer & Voyer, 2014; Falch & Naper, 2013). This paradox has sparked growing interest in the Gender Grading Gap (GGG), defined as the advantage girls receive in teacher-assigned grades compared to boys with similar performance on external assessments (Calsamiglia & Loviglio, 2019; Gortázar, Martínez de Lafuente & Vega-Bayo, 2022). Two main theoretical frameworks have been proposed to explain the GGG. The first points to gender differences in non-cognitive and socio-emotional traits: girls tend to exhibit school behaviors that align more closely with teacher expectations, such as diligence, intrinsic motivation, and self-discipline (Borghans et al., 2008; Farkas, 2003; Buchmann, DiPrete & McDaniel, 2008). These traits, although not directly captured by cognitive tests, can significantly influence grades. The second framework attributes the GGG to potential implicit biases among teachers, who may evaluate students based on gender stereotypes (Ridgeway, 2014; Gil-Hernández et al., 2024). Girls are often perceived as more competent in language subjects and better behaved in class (Homuth, Thielemann & Wenz, 2023), which may lead to systematically higher grades. The subjective nature of teacher assessment makes it particularly sensitive to contextual factors. Several studies have shown that class composition (in terms of gender, socioeconomic status, or migration background) can amplify or mitigate the GGG, due to group dynamics or teachers' cognitive pressures (Lievore & Triventi, 2023; Schuessler & Sonderskov, 2023). Moreover, the GGG tends to be more pronounced in lower-secondary education than in primary school, likely due to gender-based developmental differences and the increasing discretion of teachers in assigning grades as the educational cycle progresses (Balducci et al., 2024). Importantly, since grades strongly influence academic outcomes and educational choices, the GGG may contribute to reinforcing gendered educational trajectories, with long-term implications for inequality (Papageorge, Gershenson & Kang, 2020; Terrier, 2020; Lievore, Fedeli & Triventi, 2024).

Research Questions and Objectives

This study investigates four dimensions of the Gender Grading Gap: (1) variation across school subjects, (2) the role of school/classroom context and peer composition, (3) temporal dynamics over the school career, and (4) medium- and long-term consequences for student outcomes as measured through INVALSI standardized tests.

Data

This study relies on administrative data from the Italian National Institute for the Evaluation of the Education System (INVALSI), which offer unique advantages in terms of scope, granularity, and longitudinal depth. The data include the entire student population in Italian primary (grades 2 and 5) and lower-secondary schools (grade 8) from the 2012 to 2019 school years. In total, the dataset covers more than 7 million students across over 370,000 classrooms, allowing for analysis at a high level of granularity. INVALSI standardized tests assess students' proficiency in Italian (language literacy) and mathematics and are administered nationwide through a centralized protocol. The resulting scores, reported on a continuous and comparable scale, serve as an objective measure of academic performance, independent of teachers' subjective judgments. In addition to test scores, the data include demographic and socioeconomic variables (e.g., ESCS) and school-level context indicators, such as class composition by gender and migration background. Anonymized unique identifiers enable tracking of individual students across grades, allowing for the construction of longitudinal student panels. For the longitudinal analysis, the study follows the cohort of students who were in grade 2 in the 2012 school year and observed again in grades 5 (2015) and 8 (2018). This design allows the estimation of the GGG in the early stages of schooling and its predictive effects on later educational outcomes, capturing the cumulative impact of grading practices on student trajectories.

Method

The empirical strategy unfolds in multiple steps. The basic model estimates the Gender Grading Gap at the individual level through a simple linear regression, where the dependent variable is the teacher-assigned grade and the main independent variable is a female gender dummy, controlling for INVALSI test scores:

$$Grades_{i_s} = \alpha + \delta(female_i) + \beta(test\ INVALSI_{i_s}) + \varepsilon_{i_s}$$

The coefficient δ captures the GGG: if positive, it indicates that girls receive higher grades than boys with the same test score; if negative, it indicates the opposite. The estimation is performed separately for each subject (Italian and Mathematics) and school grade, to detect variation in the GGG across disciplines and school stages. Next, the GGG is estimated at the classroom level by clustering the regressions, allowing the analysis of heterogeneity in grading practices across educational environments. A separate equation is estimated for each class, generating a distribution of δ values that reflect intra-institutional variation. This step enables the study of how contextual characteristics (e.g., class gender ratio, average ESCS, share of students with migration background) moderate the GGG. Multilevel regression models are also employed to disentangle individual- and group-level effects, while controlling for class and school-level variables such as average test scores, internal variance, class size, gender prevalence, socioeconomic composition, and ethnic diversity. To explore temporal evolution, GGG estimates are replicated across academic years and school grades, helping to detect period trends and potential discontinuities linked to policy reforms or institutional changes in assessment. This dynamic estimation allows researchers to determine whether the GGG is stable, increasing, or decreasing over time. Finally, the study adopts a longitudinal design to examine whether early GGG values (measured in grade 2) predict later academic performance (grade 5 and 8 INVALSI scores). This approach tests the hypothesis that teacher-assigned grade differences not explained by test performance have long-term effects on student learning. Pre-treatment covariates—including prior achievement and socio-demographic characteristics—are controlled to minimize endogeneity risks.

Results

Descriptive evidence from this study indicates that, in the Italian context, the Gender Grading Gap is larger in mathematics. This may reflect a compensatory grading practice by teachers seeking to balance girls' relative underperformance in standardized math tests. Moreover, the GGG is significantly associated with class- and school-level socioeconomic and ethnic composition. The gap increases during the transition from primary to lower-secondary education, consistent with developmental and institutional explanations. The longitudinal analysis reveals that early grading advantages predict future academic success: girls who receive higher grades than boys with the same test scores are more likely to attain better learning outcomes, experience lower risk of grade retention, and enroll in more favorable educational tracks. These findings suggest that discretionary teacher grading can have lasting and cumulative effects on students' educational pathways. This study contributes to the broader debate on grading fairness and educational inequality by providing a systematic and theory-informed analysis of the Gender Grading Gap. The results highlight the need to critically reassess the role of subjective grading in reproducing gender disparities, especially when it diverges from standardized assessments of ability.

Keywords: Student achievement, grading, gender differences, grading gaps

Exploring educational inequalities through multilevel intersectional analysis (MAIHDA)

Leonardo Grilli – Enrico Contin

Introduction

Understanding disparities in student performance requires more than accounting for isolated background characteristics: it requires exploring how multiple identity dimensions intersect. Educational outcomes are influenced by an intricate interplay of gender, origin, family dynamics, and socioeconomic status. To account for the complex intersectional reality, in this paper, we adopt the Multilevel Analysis of Individual Heterogeneity and Discriminatory Accuracy (MAIHDA) framework, a method rooted in social epidemiology and increasingly applied in broader social sciences to capture nuanced inequalities (Prior et al., 2024; Evans et al., 2024).

Although MAIHDA has proven valuable in fields like health and gender inequality, its use in educational research remains limited (Keller et al., 2023; Prior and Leckie, 2024). This study contributes to filling that gap by applying MAIHDA to assess performance differentials among Italian fifth-grade students. The goal is to identify intersectional profiles most associated with extremely high or low achievement, providing evidence for targeted educational interventions.

Data and Context

The research relies on data from INVALSI, Italy's national evaluation agency. Specifically, we exploit the 2022/2023 sample for grade 5. The dataset comprises 16,011 students across 501 schools, each accompanied by rich background and socioeconomic data. Here we focus on the determinants of the mathematics score, but the approach can be applied to all available outcomes.

The main outcome variable is the WLE (Weighted Likelihood Estimate) mathematics score based on the Rasch model. The national mean and standard deviation are 191.82 and 41.07, respectively.

To define intersectional strata, seven key variables were selected:

- Gender (male or female)
- Student origin (native, first-generation immigrant, second-generation immigrant)
- Family environment (a composite score based on siblings, books, and access to educational resources)
- Parental education (each parent separately, with four levels)
- Parental occupation (each parent separately, with five levels)

The family environment was obtained by a Graded Response IRT model, incorporating the number of siblings, books at home, and the availability of study resources (e.g., dictionaries, tablets, quiet space).

Except for gender, all the considered variables are affected by missing values, posing a serious methodological challenge. We adopted different solutions for the variables, as outlined below.

The origin variable has a low rate of missing values (4.86%); therefore, students with unknown origin were dropped.

The indicators of the family environment have missing values ranging from 1.69% to 5.31%. As we summarise those indicators by an IRT model, a score is produced also for students with partially missing items.

Parental education and occupation are unknown for about a quarter of the students, with a large overlap between the cases of missing value for education and occupation. The missingness is likely to be informative; thus, instead of excluding these observations, we added an extra category to each variable. Since family background information is collected directly by school offices, nonresponse may be due to specific reasons, such as parents' perception of the questions as sensitive or a desire to avoid social judgment (Groves et al., 2011). The addition of an extra category accounts for the peculiar family background of students with missing information on parental education and occupation, with no need for assumptions about the missing data mechanism. The drawback of this approach is the relevant increase in the number of potential intersectional strata.

Combining all variables yields 21,600 theoretical intersectional strata, of which 3,362 are represented in the actual data.

Descriptive analysis revealed a largely balanced gender distribution and a student population that was predominantly native-born (87.85%), with small proportions of first- and second-generation immigrant students. The granular categorisation of parental background and home environment sets the stage for a rich intersectional analysis.

Methodological Framework

To identify student profiles associated with the lowest and highest test scores, we estimate two cross-classified multilevel linear regression models with random effects for both the intersectional strata and the schools. Students (level 1 units) are simultaneously nested within intersectional strata (first type of level 2 units) and schools (second type of level 2 units).

First, Model 1 is a random intercept linear model with no covariates. The main goal of this unadjusted model is to identify the strata with the most extreme scores. To this end, the strata are ranked based on their random effects, which are predicted with the Empirical Bayes method. Letting y_{ijk} be the math score of student *i*, belonging to stratum *j* and school *k*, Model 1 is specified as follows:

$$y_{ijk} = \alpha + u_j^{stratum} + v_k^{school} + e_{ijk} \quad (1)$$

where α represents the overall mean score, $u_j^{stratum}$ and v_k^{school} are the random effects for stratum *j* and school *k*, respectively, and e_{ijk} is an individual-level error. All random components are assumed to be independent and identically distributed with a normal distribution with zero mean and variances σ_u^2 , σ_v^2 and σ_e^2, respectively. A key part of the analysis is the calculation of the variance partitioning coefficient (VPC) for both the strata and the schools to quantify the proportion of variance of the outcome attributable to each level.

In the second model (Model 2), we add the socio-demographic variables used to define the intersectional strata as covariates:

$$y_{ijk} = \alpha + \boldsymbol{x}'_{ijk}\boldsymbol{\beta} + u_j^{stratum} + v_k^{school} + e_{ijk} \quad (2)$$

where $\boldsymbol{x}_{ijk}$ is the vector of the variables defining the strata, and $\boldsymbol{\beta}$ is the vector of regression coefficients. The purpose of this adjusted model is to disentangle the intersectional effects of Model 1 into main effects, summarised by the regression coefficients $\boldsymbol{\beta}$, and interactive effects, captured by the strata random effects. The interactive effects arise when the relationship between the outcome and the covariates is not purely additive but depends on their joint values. For example, gender differences depend on the values of origin, family environment, and parents' educational level and occupation. The role of interactive effects, which are at the core of the intersectional analysis, is summarised by the proportional change in variance (PCV) comparing the variance of the intersectional strata between Model 1 and Model 2.

The analysis is conducted with the mixed command of Stata 18.

Model results

In terms of variance partitioning coefficients (VPC), the findings from Model 1 highlight that 11.75% of the total variance in scores lies between strata, and 8.95% between schools; thus, intersectional differences account for more variation in performance than school affiliation.

Introducing the covariates in Model 2 drastically reduces the variance between strata by 97.27% (proportional change in variance, or PCV). This suggests that most intersectional variation is due to additive main effects. Nonetheless, the remaining 2.73% signals the presence of non-additive, interactive effects – true intersectionality in action.

The regression coefficients indicate strong effects:

- Males outperform females by 8.5 points on average.
- First-generation immigrant students underperform native students by 14.54 points.
- Family environment positively correlates with math scores, confirming the relevance of home conditions in educational outcomes.

After fitting Model 1 (without explanatory variables), the strata random effects have been predicted by the Empirical Bayes method. This allowed us to identify the extreme strata, namely the student profiles associated with top and bottom performance at the math test. It is worth to note that the ordering of the strata by raw average score differs from the ordering by Empirical Bayes prediction, as the latter accounts for the reliability by applying a shrinkage toward the zero inversely related to the stratum size.

The highest scores are among native male students with high family environment, whose parents have higher education levels and stable employment. In contrast, the lowest-performing strata predominantly consist of female students, either native or foreign-born, with low family environment and parents with limited education.

Final remarks

This research demonstrates that nuanced socio-demographic profiles explain a meaningful proportion of educational disparities. The MAIHDA approach allows analysts to identify not just average trends, but the profiles of students with exceptionally high and low results.

Intersectional stratification provides insight into how combined vulnerabilities, such as being a second-generation immigrant girl from a low-resource household, amplify disadvantage in ways not visible when studying variables in isolation. The findings can be used to inform policy decisions, particularly those aimed at equity and targeted support for vulnerable groups

Missing values have been handled by defining a further category for each variable. This approach is practical and effective, but it increases the number of strata. Alternatively, missing values could be filled in by a multiple imputation method.

The analysis can be developed by examining how the extreme profiles are distributed across geographic areas in Italy. Furthermore, it is of interest to apply the multilevel intersectional approach to other educational outcomes of the same sample of students, such as the reading test score, and compare the extreme profiles across outcomes to assess the extent to which they overlap.

Keywords: Interactions, random effects, schools, test scores

AUTHORS

Giovanni Abbiati is associate professor in economic Sociology at the University of Brescia. His research is focussed mainly on public policy evaluation and educational inequalities.

Tommaso Agasisti is a Full Professor at Politecnico di Milano, where he focuses on public management and performance evaluation in education and public institutions. He is Associate Dean at the PoliMi Graduate School of Management and chairs the Scientific Committee for the National Digital School Plan at the Italian Ministry of Education.

Carmen Aina associate professor at the Università del Piemonte Orientale (UPO), Delegate UPO for the Italian University for Sustainable Development, Secretary of the Italian Labour Economics Association, and Global Labor Organization (GLO) affiliate. Her research field is empirical labor economics, with particular interest in education economics, labor market transitions and early labor market outcomes.

Graziella Arazzi born in Savona on 19.12.57, former teacher t. i. class A019 - Philosophy and History at the Liceo Scienze Umane e Artistico in Imperia, currently collaborator of the USR Liguria. He collaborates with INDIRE (Training of newly hired teachers, Small Schools Movement) and INVALSI (Area 3 Evaluation of Schools; Area 5 Innovation and development). Latest publications: *School and territories: plots of coevolution to reinvent the world*, in "IUL Research", vol. 3, n. 5 (2022), pp. 230-239.

Gianluca Argentin is an associate professor in sociology at the University of Milano-Bicocca. His research interests focus on educational inequalities, teachers, and public policy evaluation.

Giulia Assirelli is a researcher at the Department of Sociology of the Università Cattolica del Sacro Cuore. She works on quantitative methods for social research and on the evaluation of the impact of public policies in the educational field.

Raffaele Auteri is currently enrolled in the LM-88 "Sociology of Networks, Information and Innovation" course and is interning at the LaPoss Research Center of DSPS UniCT. He is a L-40 Sociology graduate from the University of Catania.

Stefano Babini teaches mathematics and physics. He is passionate about problem solving, didactic communication and new technologies applied to teaching. He deals with learning processes and assessment in various educational and system contexts. He is part of the research group in mathematics didactics "Divertical-Math". He has been collaborating for years with INVALSI. He collaborates with the University of Parma.

Silvia Bacci PhD, is Associate Professor of Statistics at the University of Florence. Her research interests are methodological developments in the fields of latent variable models with a special focus on latent class models, multilevel models, item response theory models, and models for longitudinal or panel data; applications are mainly oriented to the educational field.

Francesco Balzaretti is a PhD student in Sociology and Methodology of Social Research at the University of Milan. He is part of the EDUlab research group, which deals with the analysis of educational policies and their impact on educational outcomes. His research interests include the role of institutions in combining efficiency and reducing inequality.

Paolo Barabanti is a researcher at INVALSI, where he works on teacher training, data analysis, and dissemination of national survey results, as well as research on students' digital competences. He also teaches at university level on learning assessment topics and serves as an expert evaluator for the European Commission. His research interests include gifted students, educational inequalities, data literacy, multicultural schools, learning assessment, and school dropout.

Eugenia Bellini is a research associate at Istat. For many years, she has been working on immigration issues, focusing on the implementation of sample surveys and the analysis of administrative data. She is an

expert in record linkage techniques, which she has applied to numerous and diverse datasets used in various Istat projects and scientific publications.

Piera Bello is an Associate Professor of Economics at the University of Bergamo. She holds a PhD in Economics from the Università della Svizzera italiana (USI). Before joining the University of Bergamo, she worked as a postdoctoral fellow at the University College London and at the University of Zurich. Her research has focused on gender and education economics, and household finance. Her current research agenda aims at understanding the role of teachers' stereotypes in affecting the performance and education choices of students, and at investigating the determinants of female under-representation in academia.

Pierangelo Berardi is a contract professor and coordinating tutor at the University of Foggia. He serves as an expert in TFA workshops at the University of Basilicata and in ICT-based teacher training for support education at the University of L'Aquila. He has extensive experience in teacher training within local educational districts.

Monia Berghella Secondary school teacher of Italian literature, currently serving at the Emilia-Romagna Regional School Office (USR). Expert in system evaluation, learning assessment, and teaching methodology. She has designed, coordinated, and conducted numerous professional development courses for school staff. Former evaluator for AICQ-SICEV and has held leadership roles in AICQ Education at both regional and national levels.

Maria Berrittella is an Associate Professor in Public Finance and Economics at the University of Palermo, Department of Economics, Business and Statistics. She obtained the MSc and MPhil in Environmental Economics from the University of York and the PhD in Economic Sciences from La Sapienza University, Rome. Her current research interests include government expenditure, taxation, gender economics and education.

Giulia Biazzi PhD candidate in Business & Law at the University of Brescia, with a research project focused on the evaluation of interventions to counter educational poverty. She holds a master's degree in Sociology and Social Research (University of Trento) and, since 2016, has been working as a researcher and project designer at Centro Studi Socialis, specializing in the evaluation of social interventions, particularly in the field of educational poverty.

Angela Biscaldi is Associate Professor at the Department of Social and Political Sciences, Università di Milano. Her research focuses on ethnography of communication, with particular interest in the performative, agentive and indexical aspects in everyday educational practices. Her recent publications include: "Etnografia della responsabilità educativa" (Archetipo 2013) "Antropologia dei social media. Comunicare nel mondo globale" (Carocci 2019); "Ethnography. A Theoretically Oriented Practice" (Palgrave, 2021), "La didattica a distanza. Nuove rappresentazioni dell'insegnare e dell'apprendere" (*Narrare i gruppi*, 2021), *Social media e Politiche dell'identità* (Ledizioni, 2023).

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.

Diego Boerchi University lecturer and researcher at Università Cattolica del Sacro Cuore; senior consultant for the "Accreditation and Quality of Training" Research Group at INAPP (National Institute for Public Policy Analysis). An expert in vocational guidance and career counseling psychology, he has developed and validated tools for the assessment and self-assessment of technical skills, transversal competencies, and motivations in educational, training, and work contexts.

Luca Bonacini is Assistant Professor (RTT) at the Università del Piemonte Orientale and is a research fellow at Global Labour Organization. His main research interests concern income distribution and inequality of opportunities, with a focus on education and gender issues.

Emanuela Botta is an Associate Professor of Experimental Pedagogy at the University of L'Aquila. Her research focuses on assessment, with particular attention to its equity and effectiveness, spanning from educational assessment to system-level evaluation. Part of her work is dedicated to studying educational well-being and analyzing teacher profiles. She has contributed to the development and validation of specific assessment tools, also exploring the relationships between well-being, anxiety, and assessment processes.

Massimiliano Bratti (Ph.D.) is a professor of Economics at the University of Milan. He has worked on the determinants of educational choices and educational performance, as well as the returns to education. His research has appeared in various journals, including the Journal of Public Economics, the Economics of Education Review, and the Journal of the Royal Statistical Society Series A.

Martin Brown PhD is Professor of Education at the Institute of Education, DCU (Dublin City University), and co-director of EQI – The Centre for Evaluation, Quality and Inspection. Further details can be found at: selfevaluation.eu.

Luca Bungaro holds a PhD in Statistics from the University of Bologna. His research interests focus on latent variable models, particularly IRT models and models for computerized adaptive testing.

Stefania Capogna PhD in "Social Systems, and public policies analysis PhD in "Social Systems e , and public policies analysis. Associate Professor at Link Campus University in Rome (Italy). Delegate of the Rector for Quality Assurance, Director of the LINK Doctoral School. Founder and Scientific Director of the "Research Centre Digital Technologies, Education & Society at Link Campus University. Founder and Scientific Director of the Quaderni di Comunità. People, Education and Welfare in Society 5.0, Eurilink University Press, Rome.

Michele Cardone holds a Degree in "Statistics for demographic and social sciences" and a Master (I level) in 'Statistics for the management of information systems' (Università di Roma "La Sapienza"). Working for INVALSI since 2004, member of the Statistical Service since 2010, mainly involved in the analysis of the school tests data and in the management of the annual data return to schools.

Francesca Carimando is a first-year PhD candidate in Sociology, at the University of Bologna. Her research focuses on educational inequalities, with particular emphasis on how educational trajectories and outcomes might be influenced by students' migration background.

Diego Carrasco, PhD is an Assistant Professor at Faculty of Education, 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.

Maria Cristina Carrisi associate Professor of Mathematical Physics, in possession of the National Scientific Qualification to Full Professor. She has always been interested in the didactics of mathematics and carries out research in this field as well, creating activities that develop mathematical skills through the teaching of Informatics or the use of technological tools. In addition, she studies the impact of generative AI on the teaching/learning process of mathematics.

Giorgia Casalone is Associate Professor of Economics at the University of Piemonte Orientale and a Research Fellow at the Dondena Centre, Bocconi University. Her research focuses on education and public policy.

Juan-Carlos Castillo is associate professor at the Department of Sociology, Universidad de Chile, principal researcher at the Center for the Study of Conflict and Social Cohesion (COES) and at the Millennium Nucleus on Digital Inequalities and Opportunities (NUDOS). He obtained his PhD in Sociology from the Humboldt Universität zu Berlin. Additional information: https://jc-castillo.com/.

Maria Concetta Carruba Associate Professor PAED-02, Didactics and Special Pedagogy. She teaches Special Didactics, Special Pedagogy II, and Design for Inclusive Communication. She has always been interested in learning technologies and in promoting coexistence, social equity, and inclusion. She serves as Inclusion Delegate for the Faculty of Human, Educational, and Sports Sciences; Scientific Director of the first-level

Master's program "The Use of Technologies for Inclusive Teaching (Special Educational Needs and Specific Learning Disorders)" and of the second-level Master's program "School Pedagogy: Interventions in Inclusive Teaching and Learning Difficulties.

Elvira Celardi Ph.D., is Assistant professor at the Department of Political and Social Sciences at the University of Catania in Italy. As of January 2022, she teaches social research methodology and general sociology. She primarily focuses on assessing projects that promote economic, social, and cultural development in the most impoverished and marginalized areas of Southern Italy.

Zhijun Chen is a PhD researcher at the University of Bath. Her research focuses on teaching effectiveness and inequality in education, using large-scale international assessment data such as TIMSS and TALIS. She is particularly interested in how instructional practices and classroom climate shape student learning across diverse systems.

Manuel Cheyre I am a social psychologist with a Master's in Public Policy and extensive experience evaluating educational programs in public policy contexts. Currently pursuing a PhD in Education, I am developing a reconceptualization of educational quality as an evaluative judgment based on axiological frameworks about what we value in education, supported by a comprehensive system of common educational criteria.

Maria Chirico support teacher in Secondary Education Institute and Instrumental Function Orientation, Orientation Tutor Teacher, Planning of Prevention/contrast School Dropout Actions, External Observer Invalsi. Designer, Trainer, Coordinator and Evaluator of professional training courses in FP institutions in Puglia.

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.

Cinzia Conti is a Senior Researcher at Istat and teaches "Migration and Development" at the University of Rome "La Sapienza." She holds a PhD in Demography and has been working for decades on immigration and integration, with particular attention to second-generation immigrants. She is the author of numerous publications, including *"Nuovi cittadini. Diventare Italiani nell'era della globalizzazione"* (New Citizens. Becoming Italian in the Era of Globalization), co-authored with S. Strozza and Tucci, published by Il Mulino in 2021.

Enrico Contin obtained the MSc in Statistics and Data Science at the University of Florence with a thesis on multilevel intersectional analysis to identify extreme profiles in student test scores.

Gaetano Francesco Coppeta is a Postdoctoral Research Fellow at the Department of Management, Economics, and Industrial Engineering at Politecnico di Milano. His research focuses on efficiency and productivity analysis, with a particular interest in university performance evaluation and education policy. He is also interested in individual well-being and higher education governance.no

Marta Cordini assistant Professor in Urban Sociology, earned her Ph.D. in 2012 from the Università Cattolica del Sacro Cuore in Milan. Her research focuses on housing, educational inequalities, and school and residential segregation. She has held visiting scholar positions at Rutgers University (NY) and the Autònoma University of Barcelona at the GEPS Research Centre. Her book, "Scuole e Territori: La Segregazione Scolastica a Milano", co-authored with Andrea Parma, was published by Il Mulino.

Annalisa Cristini is Full Professor of Economics at the University of Bergamo where she is Deputy Rector for Welfare and Sustainable Development. She holds M.Phil. and D.Phil in Economics at the University of Oxford. Her research interests range from macroeconomics to labour market issues. Recently she has addressed gender inequality in relation to implicit stereotypes in primary schools and university performance during COVID. From 2015 to 2021 she directed CESC, the University center for socio-economic dynamics and cooperation; she held the positions of Department Director, President of Master Course,

member of the Academic Senate and of the Board of Directors of the University of Bergamo; for the three-year period 2018-21 she was appointed by the Bank of Italy as a member of the Commission for the assignment of Bonaldo Stringher scholarships.

Chiara Rita Cunsolo is a PhD candidate in Statistics at the University of Bologna. Her research interests lie in latent variable models for longitudinal data, with a focus on the educational field.

Althea Bianca Dal Ben - PhD student at the Department of Sociology and Social Research at the University of Milan-Bicocca. My research interests focus on the topic of educational inequalities, with respect to their intergenerational transmission. Currently, I am working on the role of the family within this dynamic.

Cristiana D'Anna is Associate Professor of Methods and Didactics of Motor Activities (SSD M-EDF/01) at the Department of Education and Sport Sciences of the Pegaso Digital University, where she serves as the scientific coordinator of the "Centre for Sport Science and Human Performance." She teaches at the School of Sport (Campania and Molise), is a national trainer, and a member of the Educational and Research Centre of the Italian Gymnastics Federation (FGI).

Federica Dalponte is PhD candidate in Intersectoral Innovation at the University of Milan and Fellow at SDA Bocconi School of Management in the Government, Health and Not for Profit Division. She holds a bachelor degree in Economics and Social Sciences and a master's degree in Economics and Management of Government and International Organizations, both from Bocconi University.

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).

Paolo Davoli Graduated in Physics, PhD, former Technical Manager (inspector) and coordinator of the Regional Inspection Service at the USR-ER. Former school principal, teacher and university lecturer. Member of the Provincial Committee for Evaluation of the Education System in the Autonomous Province of Trento. Engaged in training and support for the Ministry of Education and schools on National Evaluation System (SNV), student assessment, student assessment and examinations, digital technologies, curriculum design, teaching methodologies.

Riccardo De Alessandri holds a degree in Political Science and is currently enrolled in a master's program in Economics and Management at the University of Eastern Piedmont. Gained experience at the European Commission working on ICILS data analysis and education policy evaluation.

Alessandra De Angelis teacher of mathematics and physics, Invalsi, RAV and PdM contact in Secondary School. PhD in Energetics. Collaborates with the University of Udine in the teaching field and is part of the research group in Physics Teaching. Regional secretary for Friuli Venezia Giulia of the AIF.

Willem De Cort is a PhD researcher at the faculty of Economics & Business at KU Leuven (Belgium) with expertise on teacher shortages and the labour market choices of students.

Giuspeppe Filippo Dettori is Professor of Special Education and Pedagogy at the University of Sassari and is an honorary judge at the Juvenile Court of Sassari. He is coordinator of Laura's degree course in Education Sciences and Director of the Specialization Course for the training of support teachers.He is a member of the Italian Association of Magistrates for Minors and Families, of which he was also a member of the Board of Directors, and is a member of the Scientific Committee of the journal «Minorigiustizia».

Alessandro De Santis is a psychologist, PhD student at the University of Foggia - Department of Humanistic Studies. His research area concerns the study of emotions in children. In particular, he deals with investigating within schools related to early childhood how the ability to recognize the emotions of others, allows in the school environment to prevent phenomena of violence and future deviant behavior

Marta Desimoni PhD, is a Senior Researcher at INVALSI. She is the scientific lead of the methodological and psychometric unit for INVALSI national assessment. Her research interests include measurement in psychological and educational fields and latent variable modeling.

Kristof De Witte (PhD) is a Full Professor in education economics and political economy at the Faculty of Economics and Business at KU Leuven, Belgium, and he holds the chair in effectiveness and efficiency of educational innovations at UNU-MERIT at Maastricht University, Netherlands. At KU Leuven, he is a founding member and the director of the research group Leuven Economics of Education Research. Professor De Witte's research interests comprise education economics, performance evaluation and political economy. He has published more than 100 Web of Science publications and has a strong track record in leading and coordinating large research projects (including two Horizon Europe projects). He is the recipient of numerous awards and honours, including the 2021 Pioneer Award of KU Leuven and the 2020 Laureate of the Academy Award from the Royal Flemish Academy of Belgium for Science and the Arts.

Maria Clara Dicataldo is a PhD student in Education Sciences at the University of Foggia. Her research project aims to investigate the pedagogical approaches used in the field of Business Education, with the ultimate goal of reducing the skills gap between higher education and the labor market. In particular, her research focuses on studying the educational and training models adopted by the world's top business schools.

Raffaele Di Fuccio is Associate Professor of Special Pedagogy and Didactics at Pegaso Digital University. A biomedical engineer, he holds a PhD in Human Mind and Gender Studies from the University of Naples Federico II. His research focuses on technologies for multisensory learning, tangible user interfaces (TUI), serious games, and mixed reality.

Anna Dipace is Full Professor of Experimental Pedagogy at Pegaso Digital University, where she serves as Dean of the Faculty of Human Sciences, Education and Sport. She has coordinated national and international projects on blended learning and educational innovation in school and university contexts.

Anna Rita Dipierro is a Research Fellow in Econometrics at the Department of Economics, Statistics and Finance of the University of Calabria. With a PhD in Economics and Management of Sustainability and Innovation, her research activity focuses on efficiency studies and machine learning applications in the public and private sectors. She is the author of articles published in international indexed and peer-reviewed journals.

Marilena di Padova primary school teacher, secondary school teacher for humanistic-literary disciplines, specialized in student support activities and teacher of Italian as a second language (L2). She's a PhD student in Learning Sciences and Digital Technologies at the University of Foggia. Her main research interest regards technology-enhanced learning to implement effective teaching with active methodologies, Medical Humanities, AI, educational research and inclusion.

Francesca Dota is a researcher at Istat and holds a PhD in Methodology of Social Sciences. She has extensive experience in the field of social surveys at the National Institute of Statistics, particularly on topics related to the social integration of foreign citizens and the transition to adulthood of young people with a migrant background. She has contributed to various research projects and authored numerous publications.

Emanuela Dreassi PhD, is Full Professor of Statistics at the University of Florence. Her research interests are methodological developments in hierarchical Bayesian models with spatial dependence, knockoffs constructions for variable selection, and predictive inference; applications are mainly oriented to the biomedical field.

Silvia Duranti is a researcher at the Istituto Regionale per la Programmazione Economica della Toscana (Irpet) which supports Tuscany Region in developing and evaluating public policies. Her main areas of interests are the analysis of the education system and the evaluation of the effectiveness of labour and education policies.

Graziana Epifani PhD in Social Psychology, is a Researcher at INVALSI in the Innovation and Development Area. She contributed to the design of the Italian National Evaluation System and is involved in the Erasmus+ QUALAS and PRIN projects. Her research focuses on education policy evaluation, Learning Analytics, and School-to-Work Programs as tools for student guidance.

Sofia Ercolanoni – PhD candidate at the Department of Sociology and Social Research, University of Milano-Bicocca. My research interests focus on the concept of educational poverty and digital competence, with a particular emphasis on early digital exposure and digital education. I'm currently involved in the INVALSI project "DIGCOMP.MISS".

Nurullah Eryilmaz is a Senior Research Analyst at IEA Hamburg. His work focuses on international large-scale assessments, educational equity, and advanced quantitative methods in education research.

Laura Evangelista Coordinator of the National Reference Point for EQAVET in Italy; Head of the "Accreditation and Quality of Training" Research Group at INAPP (National Institute for Public Policy Analysis). She leads research projects on the monitoring and evaluation of training accreditation systems and the implementation of quality assurance mechanisms in line with principles and indicators defined at the European level.

Rosa Fabbricatore is Junior Assistant Professor at the University of Naples Federico II. Her research focuses on latent variable models, with a particular emphasis on their innovative application within the field of Learning Analytics. She is dedicated to developing and applying statistical methods to several areas, including social sciences, educational psychology, and social health psychology.

Francesco Facciorusso is a PhD student in Teaching and Learning Sciences at the University of Macerata. His main interests include technological innovation for teacher training and the use of artificial intelligence in education. He develops educational materials for teachers, focusing on practical applications and digital tools to enhance inclusive teaching and learning experiences.

Patrizia Falzetti is the Head of the Evaluation Research Sector at INVALSI.She is responsible for the Statistical Office for SISTAN and for the INVALSI Statistical Service, which oversees the acquisition, analysis, and dissemination of data related to national and international assessments of student learning, both to individual schools and to the Ministry of Education. She is involved in making these data available for scientific research or dissemination purposes by providing access to microdata and open data.

Emanuele Fedeli is a type B postdoctoral research fellow at the Department of Social and Political Sciences, University of Milan. He earned his PhD in Social Sciences from the University of Trento in 2020. During his doctoral studies at Trento, he conducted analyses on social class inequalities, peer networks and the influence of relational dynamics in schools on students' academic motivation, academic skills, educational choices and behaviours. His research focuses on educational inequalities, the effect of linguistic diversity among students and the relationship between the socio-demographic composition of the student body and academic performance.

Adrien Fernandez is a statistician at the Office for the Study of Student Assessment and the data manager for France in the 2023 ICILS international survey cycle.

Francisco Ferraioli is Research Fellow at the Department of Architecture and Urban Studies of the Polytechnic of Milan. He obtained his Ph.D. in Political Science from the Universitat Autònoma de Barcelona. His research centers on the performance of public services and public administration. His work predominantly employs a quantitative methodological approach.

Francesca Ferrara is Associate Professor in Mathematics Education at the University of Torino. She is author of several research articles on national and international volumes and journals. She collaborates with INVALSI for the learning evaluation. She has been a member of the Scientific Committee of the European Society for Research in Mathematics Education in the years 2019-2025.

Simone Ferro is a researcher at the University of Milan. His research focuses on education economics, environmental economics, labor economics, and political economy.

Sara Fiasconaro is a Master's student in Sociology and Social Research at the University of Trento, she collaborates on LEARN and GESI projects and with the EDULAB laboratory of the University of Milan. She has been a visiting research student at Trinity College Dublin. Her research interests include social inequalities and stratification, sociology of education, migrations, and quantitative methods.

Sibilla Filippi head of education office, Oxfam Italia. Latest publications: Filippi S. (and L. Luatti), *Le dimensioni del ritardo scolastico. Un percorso di analisi School data based dei fenomeni dispersivi*, in *La scuola che legge sé stessa. Il valore dei dati e della ricerca educativa per l'analisi e il contrasto della dispersione scolastica*, Fondazione per la scuola, Bologna, il Mulino, 2025 (in press).

Brunella Fiore is Associate Professor of Sociology of Education and Family at the University of Milano-Bicocca. Among her publications are *To Each Subject Its Outcome: Digital Classrooms vs. Traditional Teaching – A Comparison of Learning Results* (2024), *Media Education*, and, with P. Barabanti and A. Cutro (2025), *High-Potential Inclusion: Toward an Organizational and Educational Model for Giftedness*, Sanoma.

Sebastiana Fisicaro latin and Greek teacher with extensive experience in training, professional networks, and leadership roles. Former Technical Inspector (D.T.), expert in methodological-didactic and evaluative planning. NEV coordinator, involved in DS evaluation, SNV, and OECD PISA. INDIRE trainer, CLIL-TFA at the University of Catania, certified in APPLE Professional Development.

Concetta Fonzo Deputy Coordinator of the National Reference Point for EQAVET in Italy and Deputy Coordinator of ReferNet Italy; Member of the "Accreditation and Quality of Training" Research Group at INAPP (National Institute for Public Policy Analysis). For several years, she has been engaged in the study of education systems, lifelong guidance policies, and quality assurance measures in training, both in Italy and across Europe.

Julian Fraillon is a Senior Project Advisor to the IEA. He is the international study director of the IEA International Computer and Information Literacy Study and has been the international study director of ICILS since its inception in 2013. His primary research interests are digital literacy education, assessment design, and civic and citizenship education.

Letizia Gambi is a PhD researcher at KU Leuven. She holds a BSc in Economics from the University of Bologna, an MSc in Economics from the Catholic University of Milan, and a Master's degree in Applied Social Research from Trinity College Dublin. With diverse international experience as an economic analyst, her most recent position before beginning her PhD was as an Economic and Policy Analyst for the Irish government, where she contributed to the evaluation of newly funded reforms. Her research interests focus on education economics and policy evaluation. At KU Leuven, she has actively contributed to several publications, utilizing Flemish data to address critical educational issues. Her work includes investigating the effects of Covid-19 school closures on student outcomes and examining the implications of teacher shortages in primary schools.

Marina García-Herrero doctoral Student with a grant from the Ministry of Science, Innovation and Universities in the area of Research Methods and Diagnosis in Education at the University of Salamanca. Member of the Consolidated Research Unit GRIAL, she belongs to the Research Institute on Educational Sciences (IUCE). Her lines of research are related to large-scale assessments, educational equity and academic resilience.

Francisco Gatica is a Research Associate at the Department of Education at the University of Bath. After previous work at the UNSECO regional office in Santiago, he currently works as Project Researcher for the International Civic and Citizenship Education Study (ICCS) 2027.

Daniela Ghio is a tenure-track associate professor at the University of Catania. She earned her PhD in Demography from the Université de Montréal (Canada) and specializes in migration and integration policy.

She studied statistics, economics, law, and political sciences at the University of Pisa. Daniela adopts demographic methods and econometric approaches to analyze the social and human capital determinants of migration and assess the impacts of spatial patterns on population dynamics.

Paola Giangiacomo She is a Senior Researcher at INVALSI in the Statistics Area, where she is Scientific lead for the DIGCOMP research program. In this role, she is responsible for measuring digital competence at the end of compulsory education. She also serves as National Data Manager for international surveys promoted by the OECD. Her main activities include the revision and adaptation of assessment instruments, the definition of sampling plans, statistical analysis of quantitative and qualitative data on large national and international databases, the drafting of technical-scientific reports, and training activities for data analysis. In her institutional role, she actively promotes access to INVALSI data, making it available for scientific research and dissemination.

Paola Giannoni Lecturer at the Department of Education Sciences (DiSFor),– UNIGE. Post-doctoral research fellow at the Department of Education Sciences - University of Genoa, from March 2022 to date. Member of the Interdepartmental Center on Inequalities and Social Policies, DiSFor (Department of Education Sciences) – DiSPI (Department of Political and International Sciences), University of Genoa.

Marco Giganti holds a PhD in Human and 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 and standardized assessment), digital competences and their assessment, and teacher training (pre-service and in-service).

Carlos J. Gil-Hernández is an assistant professor in Demography at the University of Florence and holds a PhD in Sociology at the European University Institute. His research focuses on social stratification and demography.

Tommaso Giorgi is a PhD candidate in Statistics at the University of Bologna. He works on multidimensional IRT models and, during his master's thesis, explored supervised classification methods in the analysis of digital competencies.

Ivan Graziani teaches mathematics and science. Trainer in mathematics didactics. Expert in educational design and assessment. Passionate about problem solving and didactic communication. He is a member of the "Gruppo di Ricerca e Sperimentazione in Didattica della Matematica - Pisa" (GRSDM) and the "Divertical-Math" research group. He has been collaborating for years with UNIBO, INVALSI and Mondadori-Rizzoli educational.

Grazia Graziosi holds a PhD and is a Researcher at INVALSI in the Innovation and Development Area. She has contributed to PRIN and Erasmus+ QUALAS projects. A former visiting researcher at KU Leuven, she has published internationally on education economics, school efficiency, and public policy evaluation, with a focus on equity, professional integration, and the use of counterfactual methods.

Serena Greco is a researcher at INDIRE. She conducts research on school governance models, the role of leadership and middle management, and strategies for promoting and supporting effective innovation processes. She is a lecturer at the IUL Telematic University.

Leonardo Grilli is full professor of Statistics at the Department of Statistics, Computer Science, Applications "G. Parenti" (DiSIA) of the University of Florence. He teaches courses in statistical methodology and statistical models. His main research interests concern methodological aspects related to multilevel regression models and methods for causal inference and effectiveness evaluation, with applications in the economic, social, demographic and medical fields.

Maria-Magdalena Isac is a Senior Research Fellow at the University of Bath, UK, and Director of the International Civic and Citizenship Education Study (ICCS) 2027. Her work focuses on the comparative evaluation of school systems, with an emphasis on understanding how different formal and informal

educational approaches and teaching practices contribute to young people's citizenship learning and citizenship competences.

Majlinda Joxhe is a Research Fellow at the University of Bologna and a Research Fellow at the Global Labour Organization (GLO). Her research interests concern applied and experimental economics, with a focus on international migration, gender disparities, and labor market outcomes.

Margot Joris PhD is a postdoctoral researcher at Vrije Universiteit Brussel (VUB). Further details can be found at the following link.

Alec I. Kennedy is a researcher at the IEA's Research and Analysis (RandA) Unit. Prior to joining the IEA, he worked as a policy analyst for the San Francisco Unified School District in the United States. His interests are in education policy and quantitative research methodology.

Daniel Kreisman is Associate Professor of Economics at Georgia State University, and affiliated faculty member at the University of Milan. His research is primarily in labor economics with a focus on education policy. He is also founder of CTEx - a consortium of researchers and state agencies working together to inform the future of Career and Technical Education policy in the US. Dan has a PhD in Public Policy from the University of Chicago's Harris School, and was a postdoctoral fellow at the University of Michigan. He really wanted to be a philosopher, but wasn't good enough. This made him think a lot about absolute and comparative advantage so now he is an economist.

Marika Lamacchia is a second-year PhD student in Teaching and Learning Sciences at the University of Macerata. Her research focuses on evaluating the effectiveness of digital educational content and its impact on self-regulation and soft skills in the context of sustainability. She is currently a visiting student at SUPSI in Locarno, with interests in educational neuroscience, technology, and special pedagogy.

Giovanni Lari is a 24-year-old student who graduated in Economics and Business from the University of Genoa. He is currently undertaking an internship at the European Commission (DG EAC), Unit A4 – Evidence-Based Policy and Evaluation. He is enrolled in the second year of the Master's degree program in Economics and Management at the University of Eastern Piedmont.

Francesca La Russa holds a Ph.D. in Foreign Language Teaching from Bordeaux Montaigne University and Roma Tre University. She is a researcher at INVALSI, where she works on the development of standardized tests in English as a foreign language. She is also an Associate Member of the CLLE Laboratory at Toulouse Jean Jaurès University.

Ilaria Lievore is a researcher at Fondazione per la Scuola (Turin), and she works in different projects focusing on educational inequalities, among which the Horizon project LINEup. She worked as a post-doc at UNED (Madrid), and she holds a Ph.D. in Sociology and Social Research at the University of Trento. Her main research interests include educational stratification, inequalities in educational outcomes, teacher bias, and the measurement of cognitive and non-cognitive skills, with an analytical approach focusing on micro-level mechanisms.

Pierpaolo Limone is Rector of the Pegaso Telematic University and full professor of Developmental and Educational Psychology. Former Rector of the University of Foggia (2019–2022), he is an expert in educational innovation and digital learning environments. He founded the Center for Advanced Studies in Education (CASE) and co-founded EduOpen. His scientific activity focuses on edutainment, educational technologies and teacher training, with a vast scientific production and international collaborations.

Enrico Lippo Ph.D., is currently a postdoctoral researcher at the University of Milan. His work focuses on tertiary education choices, academic performance, educational inequalities, and public policy evaluation. His research interests also include health economics, with a focus on the impact of extreme temperatures and air pollution on mortality.

Anna Loi – PhD candidate at the Department of Sociology and Social Research, University of Milano-Bicocca. My research interests focus on educational inequalities, with particular attention to students with a migrant

background. I am currently involved in a research project funded by INVALSI titled "A tool for measuring digital competence to study inequalities and educational poverty."

Andrea Lopes (corresponding author) is a second-year doctoral student at the University of Milan – La Statale. His interests include the Economics of Education, with a focus on higher-ability students, tertiary education, and academic performances. Moreover, he is also interested in Labor Economics, studying the Italian labor market.

Natalia López-Hornickel is a PhD candidate in Advanced Quantitative Methods at the University of Bath (SWDTP/ESRC) and a Research Associate at University of Roehampton. Her work focuses on attitudes towards gender equity, using large-scale data and advanced statistical methods. She has research experience in educational projects across Latin America, Europe, and the Middle East.

Lorenzo Luatti researcher on migration processes and intercultural relations, Oxfam Italia. He published numerous monographs and articles in scientific journals: Secondo Rapporto Oxfam 2024. Il sistema scolastico della provincia di Arezzo: Dati e analisi. Focus dispersione scolastica, Arezzo (2024); I libri per le scuole italiane all'estero 1890-1943 (2023); Storia sommersa delle migrazioni italiane (2020), L'emigrazione nei libri di scuola per l'Italia e per gli italiani all'estero (2017), Adulti si nasceva (2016).

Rosaria Lumino is an associate professor in general sociology at the Department of Social Sciences of the University Federico II of Naples. She works on issues related to education and policy evaluation.

Alessandra Anna Maiorano was a research fellow (M-PED/03 Didactics and Special Pedagogy) at the Department of Human Sciences for Education "Riccardo Massa", University of Milano-Bicocca. As a PhD in Education in Contemporary Society, she investigates students' intercultural sensitivity using mixed-methods.

Elisa Manzella is a research fellow at the Department of Sociology at the University of Milano-Bicocca, on the Department of Excellence project *Open Government Data. Understanding Society through Public Administration Data.* She holds a Ph.D. in Sociology from the Catholic University of Milan, with a dissertation on the role of teachers in the reproduction of educational inequalities. Her research interests include educational and guidance-related inequalities, as well as the experimental evaluation of interventions in the field of education.

Lorenzo Maraviglia - PhD in Sociology (title obtained at the University of Pisa). In the last twelve years I have directed the statistical office of the province of Lucca and I have participated in various research projects in the social and statistical fields promoted by ISTAT and various research bodies. Actually I work as a statistical researcher at INVALSI.

Francesco Salomone Marino is a doctoral candidate at the University of Palermo, undertaking studies in Economics and Statistics. He has been visiting scholar at the University of Lancaster (UK). His research is focused on social mobility, income mobility and education.

Agostino Marottoli is currently enrolled in the LM-88 "Sociology of Networks, Information and Innovation" course and is interning at the LaPoss Research Center of DSPS UniCT. He is graduate in "Sociology of Innovation". His resume include more than 40 years of professional experience in the fields of finance, ICT, organization, and tourism.

Mirko Marras I'm an inventive scientist with a Ph.D. in Computer Science and over 8 years of experience. Under my faculty position at University of Cagliari, Italy, I research and develop technology that sustains responsible AI principles — human agency and supervision, robustness and safety, privacy, transparency, diversity, equity, fairness, societal well-being, and accountability — for the benefit of individuals and society.

Michele Marsili graduated in Statistics at Sapienza University of Rome. He worked in Business Intelligence consulting, providing software development solutions for analysis and support for company's decision making in insurance and pharmaceutical industries. Since January 2018 he has been working in the INVALSI Statistical Department.

Fernando Martínez-Abad currently Professor of the Area of Research Methods and Diagnosis in Education at the University of Salamanca. He is a member of the Consolidated Research Unit GRIAL and belongs to the Research Institute on Educational Sciences (IUCE). His lines of research are related to large-scale assessments, school educational equity and effectiveness, and factors associated with school performance and success.

Rita Marzoli is the Director of the 'Aldo Visalberghi' Library and Group Coordinator for the INVALSI English test. She holds a degree in Foreign Languages and Literature (German and English) from Sapienza University of Rome and specialized in Librarianship at the Vatican School. She is currently studying for a PhD in Human Sciences at the University of Verona. Her research interests encompass language testing, information retrieval in the social sciences, systematic review methods, research evaluation, information literacy, and bibliometrics.

Antonella Mastrogiovanni graduated in Psychology, at "Sapienza University of Rome". expert in developmental age and systemic relational theories. Senior researcher INVALSI, Head of Italian tests, deals with the construction of standardized tests, in the context of reading comprehension.

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.

Giuseppe Mazzarino is an adjunct professor in Anthropology of Communication at the Politecnico di Milano. Since 2014 he has been working on the relationship between anthropology and semiotics. In 2020 he founded Fare Ricerca Academy, an entity with which he carries out research and consulting activities. Since 2021 he has been coordinating the teaching activities of the Master in Communication Design for Business at the University of Ferrara. In 2024 he founded Pensiero Selvaggio, a third sector entity with which he carries out research activities on the relationship between anthropology, design and communication. Author of Il potere dell'ipnosi. Proposte teoriche per un'etnosemiotica (2015), Tra "etno" e "semiotica" vol.I e II (2017), Fredens havn. Pratiche dello spazio in una piccola comunità galleggiante (2020), Antropologia del design. Una relazione (ancora) da costruire (2025).

Elena Meschi is an associate professor at the University of Milan-Bicocca. Her research interests include education economics, human capital, and public policy.

Alessandra Michelangeli is Associate Professor of Economics at the University of Milan-Bicocca. Her research interests include spatial inequalities, well-being, and public policy evaluation.

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.

Francesca Milzani PhD candidate in Methods and Models for Economics and Management at the University of Brescia, with a master's degree in Economics and Data Analysis (University of Verona). Since 2021, she has been working at Centro Studi Socialis, specializing in data collection and analysis for the evaluation of public policies, educational projects, and local surveys.

Jurgena Myftiu is a Researcher at University of Bergamo. Her work focuses on inequality, transport economics, the labor market and education.

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.

Veronica Mobilio leads the Research Unit at Fondazione per la Scuola of Compagnia di San Paolo. Her research focuses on educational inequalities and evidence-based practices. Previously, she worked as Policy Officer at the European Commission's Directorate-General for Education, Youth, Sport and Culture (2018–2023), leading evidence-based initiatives on innovation in education. She has also worked as Senior Manager for Education Development at Junior Achievement Europe and as a researcher at various Italian universities. She holds a Ph.D. in Social Sciences.

Serena Monica is a first-year PhD student in Mathematics at the University of Ferrara, where she also earned her master's degree in mathematics. Her main interests include the use of large-scale assessments in mathematics to identify widespread student difficulties, with a particular focus on the presence of gender gaps in mathematics and on teacher training.

Laura Montagnoli PhD, graduated in Mathematics, is a researcher at INVALSI, where she is in charge of Mathematics tests. She is a secondary school teacher on leave and is involved in teacher training.

Iacopo Moreschini is graduated in Applied Social Sciences at La Sapienza, currently research fellow at the University of Cagliari. His work centers with quantitative analysis of social inequalities in educational choices and students performances.

Sara Mori PhD in Educational Systems and Processes Evaluation, psychologist-psychotherapist. She is a researcher at INDIRE: her main research interests concern the assessment and development of transversal skills and well-being at school.

Chiara Mussida associate Professor of Economic Policy at the Università Cattolica del Sacro Cuore (Piacenza). She received her BA and Ph.D. in Economics from Università Cattolica. She also holds a Master of Arts in Development Economics from University of Sussex (Brighton, UK). Her research interests include empirical labour economics, gender inequality, education, and microeconometrics.

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 Statistical Service Area.

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 Economic Policy at the Department of Economics of the University of Bergamo, where she is also Delegate by the Rector for relations with schools and students guidance, and Director of the Center for Higher Education and Youth Employability (HEYE). She is also President of the Evaluation Committee of the school system of Trento for the period 2025–2030. After graduating in Economics and Business at the Catholic University of the Sacred Heart in Milan, she earned an MSc in Economics from the University of Texas at Austin (USA) and a PhD in Economics from the Catholic University of the Sacred Heart in Milan. Her research interests focus on labour economics, economics of education, policy evaluation and gender economics.

Francesco Orlando Graduated in Agricultural Sciences, former Technical Manager (inspector) and coordinator of the Regional Inspection Service at the USR-ER, administrative manager of a territorial Office, school principal, teacher. He deals with technical and professional education, evaluation of schools and managers, training, state exams, support for schools and networks, education, evaluation of schools and managers, training, state exams, support for schools and networks.

Marina Orselli Headmaster IC Pegli – Genoa, school regional coordination center for the training of school personnel. He collaborates with INDIRE (Small Schools Movement; management of the "Classes on the Net" project). President of the Teacher Competition Committees (A012-A022; Childhood).

Valentina Pagani is a researcher at the University of Milano-Bicocca in the Didactics and Special Pedagogy sector (PAED-02/A). She collaborates with the School and Democratic Citizenship Lab. She has published articles on the evaluation of student-teacher relationship quality and on tools for evaluative research in education.

Carmela Paladino is a first-year PhD student in Education, Development and Learning Sciences. Her research project focuses on instructional design and assessment for social inclusion. She holds degrees in Pedagogical and Educational Design Sciences and in Modern Philology, as well as the DITALS Level II certification and a specialization in didactic support for students with disabilities (TFA).

Laura Palmerio is PhD and Senior Researcher at INVALSI, she is responsible for the International Surveys area. She coordinates the OECD PISA and TALIS, as well as the IEA TIMSS, PIRLS, and ICCS studies for Italy, and serves as the national coordinator for the Horizon Europe project IMP>ACT – Impact Assessment for Action Competence. Her main research interests include equity in education, the relationship between reading and mathematics literacy, civic and citizenship education, and information literacy.

Paolo Pancheri teacher and head of the evaluation area at the "G. Marconi" Technical Institute in Rovereto. He holds a degree and PhD in Engineering. His interests include educational innovation and the use of data for school improvement. He collaborates in provincial initiatives for monitoring and evaluating teaching quality.

Ornella Papa is Specialist in Psychological Evaluation and Researcher at INVALSI Library. She has been Subject Matter Expert at the Chair of "Introduction to Educational and Evaluative Research" at the University of Rome Tor Vergata. Her current research interests are: School libraries, Information literacy, Digital divide.

Monica Papini, graduated in Statistics, at the University of Rome "La Sapienza", she obtained the II Level Master in Data Intelligence and Decision making strategies. She currently works in the Statistical Service of INVALSI where she performs support activities for statistical analysis on large data bases of national and international surveys on learning. National Data Manager for ICILS 2018.

Giuliana Parente is a research fellow at the University of Milan. She holds a PhD in Communication, Social Research and Marketing from Sapienza, University of Rome. Her research focuses on educational inequalities and the evaluation of public policies in a cross-national perspective, with a focus on gender, educational poverty, and school-to-work transitions.

Andrea Parma assistant professor of Economic Sociology. He earned a PhD in Applied Sociology and Research Methodology at the University of Milano-Bicocca. He was a visiting scholar at University of Amsterdam and University of Antwerp. He is an expert in quantitative research techniques. He works on topics such as school and territorial segregation, economic insecurity, labour market analysis and policy evaluation. He has worked on several national and international projects. His book "Scuole e Territori. La segregazione scolastica a Milano" was published with Il Mulino.

Sara Pastore PhD in Social Research Methodology, is a research fellow at INVALSI's Area 3, working on the topic of "Data Literacy and Digital Literacy in School Evaluation." Her research interests include technology addiction, digital/data/AI literacy, and educational poverty.

Serafina Pastore is an Associate Professor in the Department Research and Humanities Innovation, University of Bari (Italy). Her research examines the complex intersections of assessment practice, teacher education and educational policy as operating within the context of school and university innovations. Her recent work focuses on teacher assessment literacy.

Giulia Gabriella Pastori is a full professor at the University of Milano-Bicocca in the Didactics and Special Pedagogy sector (PAED-02/A). She coordinates the School and Democratic Citizenship Lab and researches educational inclusion, student participation, and qualitative and quantitative research methodologies.

Fabrizio Patriarca is a Professor of Economics at the University of Modena and Reggio Emilia. His research explores topics in economic policy, environmental economics, socio-economic inequalities, and the relationship between economics and political identities.

Caterina Pavese is an economist at the Ifo Institute and a CESifo affiliate in Munich. Her work focuses on education economics.

Guendalina Peconio is PhD in Neuroscience and Education and is currently a research fellow at the Department of Humanities, University of Foggia. She is part of the Learning Sciences Institute at the same university and is involved in human capital development, entrepreneurship education and teacher training in the field of evidence-based education through involvement in various research projects.

Riccardo Pietracci is a research associate at INVALSI and the national coordinator for computer-based assessments in the field of international large-scale studies. His main research interests include civic and citizenship education, sustainability, creative and computational thinking.

Andrea Pietrolucci is a Postdoctoral researcher at the University of Milan (Italy) in the ERC Consolidator Grant EDUPOL- Education policies that work: A context-sensitive 'big data' approach (2024-28). His main research interests focus on inequalities in educational opportunities, geographical disparities, and wealth inequality.

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.

Marialuisa Pinna is a teacher specializing in educational support for students with disabilities and teaches at a secondary school. She is a researcher at the Eurispes Foundation – Institute for Political, Economic, and Social Studies. Her research focuses on issues related to education and inclusion.

Francesco Pisanu director of the Office for the Evaluation of Education Policies of the Autonomous Province of Trento. He holds a degree in Psychology and a PhD in Information Systems and Organization. His work focuses on education system evaluation, data use in education, and non-cognitive skills. He has published extensively on evaluation, student guidance, and educational innovation.

Donatella Poliandri PhD is Head of Department at INVALSI (Istituto nazionale per la valutazione del sistema educativo di istruzione e di formazione). Further details can be found at the following link

Fernando P. Ponce, PhD is an academic researcher affiliated with the Department of Psychology at Universidad Católica del Maule. His research spans Experimental Psychology, investigating mechanisms underlying human behavior, and Psychological Assessment and Psychometrics, focusing on developing and validating tools to measure psychological constructs, wording effects, and cognitive processes underlying response to tests.

Mariano Porcu is a Full Professor of Social Statistics at the University of Cagliari. His research is principally devoted to the evaluation of educational systems, with a particular emphasis on the construction of adjusted composite indicators for assessing institutional performance.

Agostino Portera Professor of Intercultural Education at the University of Verona (Italy), head of the Center for Intercultural Studies, and director of the Master's program in "Intercultural Competences and Management". He studied Psychology in Rome (La Sapienza), Ph.D in education in Freiburg (Germany), and was a Visiting Professor and speacker at many Universities (Würzburg, London, UCLA Los Angeles, OISE-Toronto, Victoria Wellington, Richmond and Harvard, USA). His latest books deal with opportunities of

intercultural intervention in schools, families, and workplaces: *Intercultural Education, Curriculum Development, Assessment and Teaching* (with M. Trevisan & Marta Milani), Routledge, London 2025*; Educazione e Pedagogia Interculturale,* Il Mulino, 2022; *Education and Intercultural Identity* (written with Z. Bauman), Routledge, 2021; *Intercultural Mediation, Counselling, and Psychotherapy in Europe* (with R. Moodley and M. Milani), 2020. ORCID: https://orcid.org/0000-0002-6356-5688; Website: https://www.dsu.univr.it/?ent=persona&id=449.

Stefania Pozio was a senior researcher at INVALSI from 2017 until her retirement in 2024. She was responsible for the national mathematics assessments. Within INVALSI, she continues to collaborate both with the national mathematics assessment division and the research area dedicated to international surveys (TIMSS and PISA), always focusing on mathematics.

Giovanni Puccetti is a postdoctoral researcher at the Institute of Science and Technologies of Information "A. Faedo". His research revolves around Natural Language Processing (NLP) and the current focus is on methods to detect machine generated content, mostly text. He is also interested in the evaluation of Large Language Models in Italian and their mathematical understanding in this language.

Carla Rampichini, PhD, is Full Professor of Statistics at the Department of Statistics, Computer Science, Applications "G. Parenti" of the University of Florence. Her research interests relate to random effects models for multilevel analysis, program evaluation, and causal inference. Her methodological work is joined with applications, often concerning the effectiveness of universities and educational system.

Costanzo Ranci full Professor of Economic Sociology. He has published several books and scientific articles on topics such as welfare, social vulnerability, and long-term care policies. He has served as a consultant for the Ministry of Social Affairs, INPS (National Social Security Institute), the Lombardy Region, and many other institutions. He was a visiting scholar at the University of California – Berkeley, the University of Barcelona, the University of Lausanne, and Harvard University. His current research focuses on the following topics: middle-class insecurity, social investment policies, long-term care policies, school segregation and inequalities.

Angela Rella is a Research Fellow in Business Economics at the Department of Economics, Management and Quantitative Methods of the University of Milan Statale. With a PhD in Economics and Management of Sustainability and Innovation, her research activity focuses on the application of efficiency measurement techniques in the public sector. She is the author of articles published in international indexed and peer-reviewed journals.

Giuseppe Rizza chief School Inspector of the Autonomous Province of Trento. He holds a degree in Political Economy and previously taught Applied Mathematics. He has conducted research at the University of Trento and has led major upper secondary schools in the Province of Trento and the Veneto Region. He earned a master's degree in Digital Innovation and Inclusion and is particularly active in the fields of European citizenship and educational innovation.

Ottavio Giulio Rizzo assistant professor in Mathematics Education at Università degli Studi di Milano. His research interests include: problems related to the teaching and learning of mathematics service courses, the education of pre-service secondary school mathematics teachers — in particular about their beliefs —, the use of technology in the teaching and learning of mathematics.

María-José Rodríguez-Conde full Professor of the Area of Research Methods and Diagnosis in Education at the University of Salamanca. She is a member of the Consolidated Research Unit GRIAL and belongs to the Research Institute on Educational Sciences (IUCE). Her lines of research are related to large-scale assessment and evaluation and quality management in higher education.

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.

Sara Romiti PhD, is a Researcher at INVALSI, where she coordinates monitoring and evaluation activities for the Erasmus+ QUALAS project. She contributed to the launch of the Italian National Evaluation System and has worked on national projects on teacher training and school evaluation. Her research focuses on teachers' assessment skills, classroom teaching practices, and comparative analysis of international evaluation models.

Francesco Rubino senior officer for educational and training experimentation at the Office for the Evaluation of Education Policies of the Autonomous Province of Trento. He works on educational policy evaluation, monitoring, and instructional innovation. He collaborates on national and international projects on skill development and data use in schools.

Francesco Sacchetti associate Professor of Social Science Methodology at the University of Urbino Carlo Bo. His research interests lie in the epistemology and methodology of social research, particularly ethnography, visual approaches, and both quantitative and qualitative tools.

Cecilia Saint Pierre PhD is Director of Online Education at University of Chile and a visiting researcher at UVA (Universidad de Valladolid).

Guido Salza is an Assistant Professor at the University of Milan. He holds a PhD in Sociology and Social Research Methodology and focuses on the institutional and informal constraints that limit equitable access to education. His most recent work has been published in the European Sociological Review and the Journal of Happiness Studies.

Felipe Sánchez-Barría is PhD in Sociology, Academic at the Institute of Sociology at the Pontificia Universidad Católica de Chile and adjunct researcher at the Millennium Nucelus on Labor Politics and Colective Familiy Life - LABOFAM, Chile. His research focuses on topics related to non-conventional political participation, primarily focusing on involvement in protests and social movements and their interactions with the political system and public opinion.

Andres Sandoval Hernandez is currently a professor at the Department of Education at the University of Bath, UK. Prior to that, he worked as Head of the Research and Analysis Unit at the International Association for the Evaluation of Educational Achievement (IEA) in Germany.

Giulia Santangelo is a researcher in the Department for Vocational Education and Training (VET) and Skills at the European Centre for the Development of Vocational Training (Cedefop), an EU agency. Her research interests include labor economics, applied micro-econometrics, and the evaluation of education and labour market policies.

Miria Savioli Sociologist, specialised in social research methods and techniques. She has been a researcher at Istat since 2000. From 2004 to 2012, she led the Survey on the Conditions of Childhood and Adolescence. Since 2011, she has worked on the Fair and Sustainable Well-being Measurement Project. Since 2023, she has led the Measures of Child Welfare project and been a member of the Commission on Educational Poverty.

Teresa Savoia is Associate Professor of Experimental Pedagogy at Pegaso Digital University, where she also serves as Director of the University Language Centre. She collaborates on national and international projects focused on educational innovation in school and university contexts, as well as language learning.

Elisa Sbalchiero is a first-year PhD student in Sociology and Social Research Methodology at the Department of Social and Political Sciences, University of Milan. Her research project focuses on inclusive education policies in Europe, with particular attention to the Italian educational context, investigating their role in reducing learning disparities among students from disadvantaged backgrounds. She studies education, educational achievement and educational inequalities based on social class and migration background. She is also interested in the school-to-work transition.

Cintia Scafa Urbaez Vilchez is a research fellow in the mathematics assessment area at INVALSI. In 2023, she earned a PhD in Mathematics Education from the University of L'Aquila with a dissertation focused on

how paper-based materials and educational video games can activate scaffolding strategies between pairs of primary school students.

Elisabetta Segre PhD in Economic Sciences and researcher at Istat since 2010. She became a senior researcher in 2023 and began working with the Data Analysis and Economic, Social and Environmental Research Unit, where she worked on integrated welfare analysis. She joined the Commission on Educational Poverty in the same year.

Mike Smet is an associate professor at the faculty of Economics & Business at KU Leuven (Belgium) with expertise on transitions between education and labour market and in and out of employment.

Mara Soncin is a researcher at Politecnico di Milano in the Department of Architecture and Urban Studies (DAStU). Her work focuses on school governance, participation, organizational innovation, and education policy. She collaborates with public institutions and schools on improvement and evaluation processes.

Francesca Storai is a researcher at Indire. She has dedicated herself to the study of innovative training models for teachers and school managers. She has conducted in-depth research on methods and tools for the development of improvement plans and educational innovation.

Rolf Strietholt is the Co-Head of the IEA's Research and Analysis Unit (RandA). He coordinates research on educational measurement and consults external researchers and international organizations. His main research interests include school and educational effectiveness research and research methodology.

Cristina Stringher senior researcher at INVALSI responsible for the ECEC sector, has been Chair in the Data Development Group of the OECD ECEC Network, currently Italian representative within this Network, has been component in the Advisory Board OECD Study on Social and Emotional Skills and invited expert by the EU Commission on: quality indicators for ECEC; key competences; Learning to Learn.

Massimo Strozza is a Senior Researcher at the National Institute of Statistics. He is an expert in statistics on university students and immigration. He is currently responsible for the Istat Survey on "Children and Adolescents: Behaviors, Attitudes, and Future Plans." Specialized in methods and techniques for social research, he has authored numerous publications on the educational paths of young people and on new generations.

Isabella Sulis is a Full Professor of Social Statistics at the University of Cagliari. Her research interests include multilevel analysis, item response theory, and measurement models aimed at developing adjusted performance indicators, with a primary focus on the evaluation of the education system.

Andrea Tinterri is associate professor in Experimental Pedagogy at Pegaso Online University. He obtained a Phd in Neurobiology at Sorbonne Université in France. He is the coordinator of the Teaching, Learning and Evaluation (TeLE) laboratory at Pegaso and a professional board game designer. His scientific interests include evidence-based education, game-based learning, and Artificial Intelligence in education.

Kateryna Tkach is a Researcher at the Gran Sasso Science Institute. Her work focuses on income inequality, and regional resilience.

Pierluigi Toma is an Associate Professor of Econometrics at the Department of Economics of the University of Salento. With a PhD in Management, Economics and Quantitative Methods, he mainly deals with efficiency and productivity analysis with the application of non-parametric methods to microeconomic issues. He is the author of books and articles published in international indexed and peer-reviewed journals, with a high Impact Factor and numerous citations.

Valeria F. Tortora is a researcher at INVALSI, she is Data Manager for Italy in international surveys promoted by the International Association for the Evaluation of Educational Achievement (IEA). PhD in Comparative Education, she focuses her scientific activity on the analysis of educational processes from an international and comparative perspective. Her most recent research focuses on the study of social

inequalities and the exploration of the socio-cultural and contextual determinants of student academic performance, with particular attention to the mechanisms of equity and inclusion in education systems.

David Torres Irribarra, PhD 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.

Giusi Antonia Toto is full professor of Didactics and Special Pedagogy at the Department of Humanities, University of Foggia. Her research interests include the analysis of the dynamics of educational communication and interaction in digital contexts, the evaluation of engagement and motivation processes in students, and the study of innovative teaching methodologies to foster the optimization of the learning experience and the development of soft skills.

Ernesto Treviño, Ed.D. from Harvard, is Full Professor at Pontificia Universidad Católica de Chile, director of CENTRE UC, and principal researcher at the Center on Educational Justice. He collaborates with international organizations and governments on education policy. His research focuses on the intersection of policy, school practices, and classroom interactions through the lens of educational justice.

Moris Triventi is Full Professor of Sociology at the University of Milan, Principal Investigator of the ERC-CoG EDUPOL ("Education Policies that Work") and co-PI on the LEARN Horizon-EU project. He has held positions at Trento and the European University Institute. His research spans social stratification, educational inequality, policy evaluation, and quantitative methods.

Cristian Usala is Ph.D. Assistant Professor (Non Tenured/RTDa) in social statistics at the University of Cagliari, Department of Social and Political Sciences. His research focuses on educational inequalities, territorial disparities, peer effects, and student mobility, with a focus on the application of discrete choice approaches.

Giulia Vaiani PhD student at the Department of Management, Economics and Industrial Engineering at Politecnico di Milano. She holds a Master's degree in Management Engineering from Politecnico di Milano. Her research interests focus on the use of digital technologies in school settings and on the quantitative analysis of educational processes. She collaborates on national and international projects investigating the impact of ICT on teaching and learning.

Loris Vergolini is Associate Professor at the University of Bologna since 2024. Previously, she was a researcher at the University of Bologna (2021-2024) and at Irvapp, Fondazione Bruno Kessler's Institute for Evaluative Research on Public Policies (2009-2021). His work focuses on policy evaluation, emphasising widening participation in higher education. His interests include educational inequalities, social cohesion, and ageing.

Jacob Van Belle is a PhD researcher at the faculty of Economics & Business at KU Leuven (Belgium) with expertise on teacher shortages.

Sara Vergallo Ph.D. candidate in the Doctorate of National Interest in Teaching & Learning Sciences with administrative headquarters in Macerata, Stem Teaching curriculum. The operational site of the doctorate is the University of Cagliari, where she is pursuing a research project in mathematics didactics. He also collaborates on some research in undergraduate teaching at the University of Milan.

Loris Vergolini is currently an Associate Professor at the University of Bologna. 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.

Elodie Vezon Persem is currently an expert in school inclusion, accessibility and innovation at the French Ministry of Education's Department of Evaluation, Forecasting and Performance. She is head of the

Accessibility, Innovation and Research unit, having previously worked as a teacher of students with all types of impairments, and as a trainer in inclusive education for teachers, trainers, headteachers etc.

Nathalie Vigna is a Postdoctoral researcher at the University of Milan (Italy) in the ERC Consolidator Grant EDUPOL- Education policies that work: A context-sensitive 'big data' approach (2024-28). Her research extends to topics such as social stratification, educational opportunities, subjective social class, and geographical disparities.

Viviana Vinci is Full Professor of Experimental Pedagogy at the University of Foggia. She is a member of the ANVUR Working Group on the Recognition and Enhancement of University Teaching Competences and of the Commission for the drafting of the National Curriculum Guidelines for Schools. She serves as the national coordinator of the eTwinning for Future Teachers (ITE) initiative and is the internship coordinator. She is the recipient of the 2021 SIRD Award and the 2014 SIPED Award.

Eleonora Vlach is Substitute Professor of Sociology (Social Stratification & Social Policy) at Goethe University Frankfurt since 2024. Previously, she was a post-doctoral researcher in Markus Gangl's team (2018-2024). She is Deputy Editor of the European Sociological Review, and her research examines educational and social inequalities by origin and migration.

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.

Nadir Zanini is currently Head of Assessment Measurement at the Cambridge University Press & Assessment, a non-teaching department of the University of Cambridge. He was previously Associate Director for Research, Data & Analytics and Head of Statistics at Ofqual (Office of Qualifications and Examinations Regulation in England), and researcher at FBK-IRVAPP (Research Institute for the Evaluation of Public Policies in Trento).

Via Ippolito Nievo, 35 – 00153 – Tel. 06941851
Mail: protocollo@INVALSI.it PEC: protocollo.INVALSI@legamail.it Website: www. INVALSI.it
C.F. 92000450582