MARZO 2026 | NUMERO 52

GLI APPROFONDIMENTI DEL SERVIZIO STATISTICO

La newsletter ufficiale del Servizio Statistico INVALSI

Gentilissimo/a,

in questo numero della newsletter *"Gli approfondimenti del Servizio Statistico"* proseguiamo lungo la stessa linea di analisi del precedente numero di Settembre 2025 dove abbiamo avviato un percorso di approfondimento dedicato alle interazioni tra le caratteristiche socio-anagrafiche degli studenti.

Background migratorio e fragilità degli apprendimenti: il ruolo delle interazioni

In tale sede, abbiamo mostrato come il divario nei risultati tra studenti in regola con il percorso scolastico e studenti in ritardo negli studi non sia costante nei differenti indirizzi scolastici – licei, istituti tecnici e istituti professionali – o in tutte le aree geografiche, ma vari in modo rilevante in base a tali dimensioni. Questa volta concentreremo l'attenzione sul **background migratorio** e sulle sue interazioni con altre condizioni di fragilità.

Per studenti "fragili" intendiamo quelli con una probabilità più elevata di abbandonare gli studi oppure di concludere il percorso scolastico senza aver acquisito competenze adeguate a partecipare pienamente alla vita sociale e lavorativa.[1]

Che cosa si intende per interazione

Nel linguaggio statistico si parla di **interazione** quando l'associazione tra una variabile e un risultato cambia a seconda del livello di un'altra variabile.

Nel nostro caso, il risultato è il livello di apprendimento, misurato attraverso i punteggi delle prove INVALSI. Sappiamo che tali risultati variano in base a una pluralità di caratteristiche dello studente: genere, regolarità del percorso, indirizzo di studio, area geografica, background socio-economico e migratorio.

[1] Nel Rapporto INVALSI viene utilizzata l'espressione "dispersione implicita" per indicare la condizione degli studenti che, giunti al termine del percorso di studi secondari, non dispongono dei livelli minimi di competenze basilari (italiano, matematica, inglese).

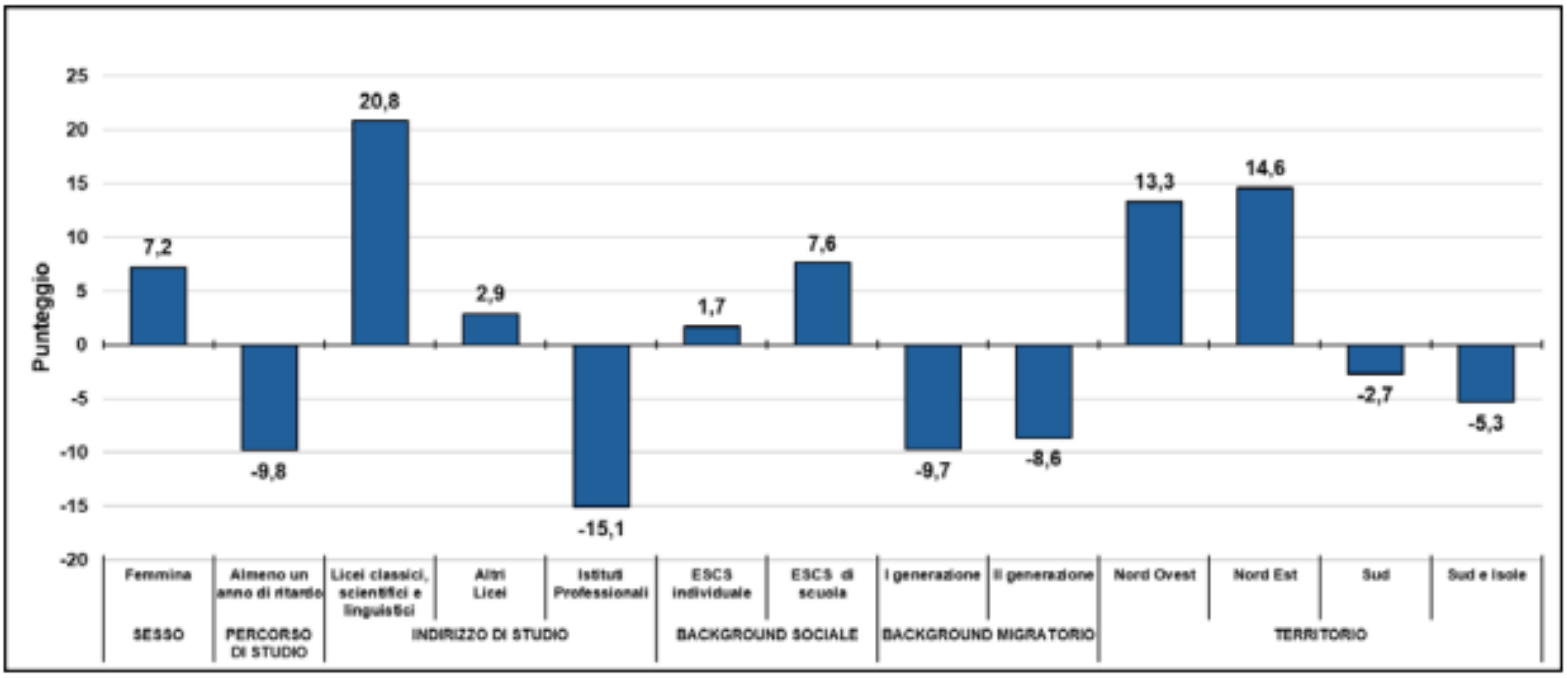

Grafico 1 – Peso di alcuni fattori sull'esito della prova di Italiano al termine del secondo ciclo d'istruzione. Differenza di punteggio (Fonte: Rapporto INVALSI 2025).

Nel grafico 1, ad esempio, è evidenziato che il punteggio medio nella prova di italiano somministrata nell'ultimo anno delle scuole secondarie di secondo grado varia rispetto a un profilo di riferimento (studente maschio, regolare, iscritto a un istituto tecnico del Centro Italia, di livello socio-economico medio). In particolare, si può notare che:

- chi frequenta un istituto professionale ottiene in media 15 punti in meno rispetto a chi frequenta un istituto tecnico;
- uno studente straniero nato all'estero (prima generazione) o in Italia (seconda generazione) da genitori stranieri ottiene circa 9-10 punti in meno rispetto a uno studente nativo.[2]

Queste differenze sono calcolate tenendo ferme tutte le altre variabili, per isolare l'associazione di ciascun aspetto con il risultato di interesse.

Ma cosa succede quando due o più condizioni di svantaggio si presentano simultaneamente? Gli svantaggi si sommano fra di loro? Oppure si rafforzano o si attenuano a vicenda?
Lo studio delle interazioni statistiche consente di dare risposte a tali interrogativi.

[2] Utilizziamo la terminologia adottata nel Rapporto INVALSI, in base a cui gli studenti sono classificati, con riferimento al background migratorio, in "nativi" (nati in Italia o all'estero da almeno un genitore italiano), di "prima generazione" (nati all'estero da genitori stranieri) e di "seconda generazione" (nati in Italia da genitori stranieri).

Un fenomeno in crescita: gli studenti con background migratorio

Il tema di come la presenza di un background migratorio interagisce con altre condizioni nell'influenzare le prospettive scolastiche degli studenti assume particolare rilievo alla luce delle trasformazioni demografiche in corso.

Secondo i dati più recenti del Ministero dell'Istruzione e del Merito[3], nelle scuole italiane sono presenti **914.860 studenti con cittadinanza non italiana**, pari all'11,2% del totale. Di questi:

- **598.745** sono nati in Italia (seconda generazione);
- **316.115** sono nati all'estero (prima generazione).

A questi vanno aggiunti circa un **10–12% di studenti con cittadinanza italiana ma con almeno un genitore nato all'estero**.[4]

Nel complesso, oggi circa **un quinto degli studenti** ha un qualche tipo di background migratorio. Questa quota è destinata con ogni probabilità a crescere nei prossimi anni.

Gli studenti stranieri, inoltre, risultano sovra-rappresentati negli indirizzi di studio con esiti di apprendimento più deboli. Nella scuola secondaria di secondo grado, infatti, essi rappresentano:

- il **5,3% nei licei**,
- il **10,4% negli istituti tecnici**,
- il **13,7% negli istituti professionali**,
- e arrivano al **28,7% nei percorsi di istruzione e formazione professionale (IeFP)**.

Poiché gli studenti degli istituti professionali presentano, in media, risultati più bassi nelle prove standardizzate sugli apprendimenti, questa concentrazione può contribuire a generare scenari particolarmente delicati.

[3] Ministero dell'Istruzione e del Merito, Ufficio di Statistica, GLI ALUNNI CON CITTADINANZA NON ITALIANA - A.S. 2022/2023 (https://www.mim.gov.it/pubblicazioni/-/asset_publisher/6Ya1FS4E4QJw/content/gli-alunni-con-cittadinanza-non-italiana-anno-scolastico-2022-2023).

[4] Nostra elaborazione sui microdati dell'indagine ISTAT Sull'integrazione delle seconde generazioni (https://www.istat.it/scheda-qualita/indagine-sullintegrazione-delle-seconde-generazioni/).

Come si analizzano le interazioni fra variabili

I dati INVALSI consentono di distinguere tre gruppi di studenti:

- Nativi: nati in Italia o all'estero con almeno un genitore nato in Italia;
- Stranieri di prima generazione: nati all'estero da genitori nati all'estero;
- Stranieri di seconda generazione: nati in Italia da genitori nati all'estero.

Per analizzare le interazioni utilizziamo modelli di regressione lineare simili a quelli presentati nella newsletter precedente, con termini che permettono agli effetti di cambiare a seconda del contesto.

In forma semplificata:

$$punteggio_i = b_0 + b_1*background_mig_i + b_2*genere_i + b_3*tipo_i + b_4*escs_i + b_5*ripete_i + b_6*area_i + b_7*(background_mig_i*genere_i) + b_8*(background_mig_i*tipo_i) + b_9*(background_mig_i*escs_i) + b_{10}*(background_mig_i*ripete_i) + b_{11}*(background_mig_i*area_i) + e_i$$

dove:

- $punteggio_i$ è il risultato nella prova (ad es. Italiano) per l'i-esimo studente;
- $background_mig_i$ è la variabile che esprime il background migratorio per l'i-esimo studente;
- $genere_i$, $escs_i$, $ripete_i$ e $area_i$ sono i valori degli altri predittori per l'i-esimo studente;
- b_0, b_1 ecc. sono i parametri del modello[5];
- e_i è la componente di errore casuale.

[5] I parametri sono gli stessi per tutti gli studenti, dal momento che (come detto) stabiliscono il peso di ciascun carattere nella determinazione del risultato atteso.

I divari medi con e senza interazioni

Molte delle interazioni del modello, ad esempio quella fra background migratorio e area geografica, sono statisticamente significative. La significatività sta ad indicare che l'importanza di una variabile sotto il profilo della previsione del risultato finale dipende in modo rilevante dalla modalità di un'altra variabile. In termini concreti, se siamo interessati a comprendere il divario fra studenti nativi e studenti di prima o di seconda generazione, sapere se stiamo confrontando fra loro giovani che vivono al Nord, al Centro o al Sud è un'informazione essenziale, di cui dovremmo tener conto nella nostra analisi.

Per chiarire questo punto facciamo qualche esempio.
Partiamo dai risultati del modello senza interazioni, quello i cui risultati in termini di scostamento atteso da un profilo medio di riferimento sono riportati nel grafico 1. Per dare qualche elemento ulteriore, specifichiamo che tale riferimento è pari a 174,4 punti. Partendo da tali informazioni, proviamo a chiederci di quanto si discosta dalla linea di riferimento la previsione per uno studente maschio, in condizione di regolarità degli studi, che frequenta un istituto tecnico e, infine, che vive in una regione del Nord Est e ha un background migratorio di prima generazione. Un soggetto con tali caratteristiche si discosta dal profilo di riferimento soltanto per le ultime due condizioni elencate, ovvero il background migratorio e l'area geografica. Per ottenere una previsione per questo studente, pertanto, dobbiamo aggiungere al riferimento (174,4 punti) l'effetto positivo derivante dal fatto di vivere nel Nord Est (+14,6 punti) e sottrarre l'effetto negativo connesso al background migratorio (- 9,7 punti); il risultato algebrico di tali operazioni (179,3) corrisponde a uno scarto positivo di circa 5 punti. E qual è, invece, il risultato atteso di uno studente "nativo" del Nord Est, sempre sotto l'ipotesi di tener ferme al profilo di riferimento tutte le altre variabili? In questo caso, dobbiamo semplicemente sommare al valore di riferimento (174,4) l'effetto positivo dell'area geografica (+14,6), ottenendo un valore atteso di 189 punti. Veniamo adesso a quello che davvero ci interessa: il divario fra studenti italiani e studenti stranieri, in questo caso specifico con un background di prima generazione. La differenza è pari a 179,3 – 189, ovvero 9,7 punti a svantaggio degli studenti stranieri. Questo è, né più né meno, il gap riportato nel grafico 1 da cui siamo partiti. Ciò non è casuale: in un modello senza interazioni la differenza di risultato fra due gruppi si mantiene costante a prescindere dal valore delle altre caratteristiche;[6] in pratica, ovunque ci spostiamo nel territorio il divario legato al background migratorio resta sempre lo stesso.

A pensarci bene, l'assunto che variabili di grande portata quali il background migratorio, l'area geografica di residenza o il percorso scolastico non presentino un qualche livello di influenza reciproca è piuttosto forte. Per questo ha senso esplorare le possibilità di modelli con interazioni.

[6] Questo risultato è vero per costruzione, dipende cioè dalle caratteristiche tecniche del modello utilizzato.

Ogni volta che si introducono nuove variabili o nuove combinazioni di variabili in un modello statistico tutti i coefficienti cambiano, a volte marginalmente a volte in modo più marcato ed interessante.

Nella tabella 1 sono riportati i nuovi coefficienti del modello in cui il background migratorio è fatto interagire con le altre variabili. Per ragioni di spazio – i coefficienti sono molti – ci limitiamo ai soli coefficienti che ci interessano più da vicino (quelli del background migratorio, dell'area geografica e delle interazioni fra tali variabili) rinviando all'appendice per la tabella completa.

	coefficiente	errore standard	significatività
Riferimento (intercetta)	173,9	0,2	
Effetti principali			
Background migratorio (rif. Nativi)			
Prima generazione	-4,1	1,1	*
Seconda generazione	-5,1	0,6	*
Area geografica (rif. Centro)			
Nord Ovest	13,7	0,2	*
Nord Est	14,9	0,2	*
Sud	-2,6	0,2	*
Sud e Isole	-5,3	0,2	*
*Interazioni (rif. Nativi * Centro)*			
Prima generazione * Nord Ovest	-4,4	1,0	*
Seconda generazione * Nord Ovest	-3,1	0,6	*
Prima generazione * Nord Est	-3,1	1,0	*
Seconda generazione * Nord Est	-1,9	0,6	*
Prima generazione * Sud	-1,0	1,2	
Seconda generazione * Sud	-3,3	0,8	*
Prima generazione * Sud e Isole	3,4	1,5	*
Seconda generazione * Sud e Isole	-2,5	0,9	*
* effetti significativi al 95%			

Tabella 1 – Coefficienti (con errori standard e livelli di significatività statistica) del modello di regressione con interazioni fra background migratorio e altre variabili (fonte: elaborazione su dati INVALSI).

In questo nuovo scenario, per ottenere la previsione per uno studente nativo del Nord Est, tenendo ferme le altre variabili ai valori di riferimento, dobbiamo aggiungere alla nuova baseline (173,9) l'effetto positivo del fatto di vivere nel Nord Est(+14,9), ottenendo un valore pari a 188,8. Per ottenere la previsione per uno studente di prima generazione dobbiamo sottrarre a tale risultato l'effetto negativo del background migratorio (- 4,1) e – ecco la novità! – il coefficiente che coglie l'effetto di interazione fra background migratorio di prima generazione e residenza nel Nord Est (-3,1). Il risultato per tale tipologia di studente è dunque 181,6. A questo punto, se andiamo a calcolare la differenza fra nativi e stranieri di prima generazione otteniamo (188,8 – 181,6) ovvero 7,2 punti a svantaggio dei secondi. Si tratta di una differenza ancora rilevante ma un po' più tenue di quella suggerita dal modello senza interazioni. Quale risultato dobbiamo dunque prendere per buono (utile, plausibile), quello del modello senza interazioni (-9,7 punti) o quello del modello con interazioni (- 7,2 punti)? Il fatto che l'interazione fra background migratorio e area geografica risulti in questo caso significativa ci fa propendere per la seconda opzione. Nel caso specifico, dunque, è un po' come se l'effetto positivo dell'area geografica (se ci si passa l'espressione "l'aria del Nord Est") avesse in qualche modo mitigato l'impatto negativo sul risultato nelle prove di italiano che tende ad esser associato alla presenza di un background migratorio di prima generazione.

Come si è detto, questa è un'interazione tutto sommato lieve anche se statisticamente significativa – e, dunque, degna di essere considerata quando si fanno analisi o previsioni. In alcuni casi le differenze sono potenzialmente più eclatanti. Nel Sud e Isole, ad esempio, ripetendo le operazioni sopra delineate con i coefficienti appropriati si ottiene una stima (a parità di genere, regolarità, percorso di studi e condizione socio-economica) di 168,9 punti per gli studenti nativi e di 167,9 punti per gli studenti di prima generazione. In questo caso, dunque, il divario è quasi scomparso. In prima battuta si potrebbe commentare tale risultato sorprendente dicendo che due condizioni di svantaggio sotto il profilo dei risultati medi nelle prove di italiano – vivere nel Sud e Isole ed essere migranti di prima generazione – anziché sommarsi tendono almeno in parte ad elidersi a vicenda. Questo ragionamento ha una sua logica e plausibilità: se gli svantaggi si cumulassero meccanicamente fra loro ci troveremmo con distanze finali abissali, ad esempio fra studenti nativi che frequentano un liceo nel Nord Est e studenti di prima generazione che frequentano un istituto professionale nel Sud. Tali divari esistono, ma non sono così esorbitanti.

Nella tabella 2 sono riportati i divari fra studenti di prima generazione e nativi e fra studenti di seconda generazione e nativi per tutte le aree del nostro Paese. Le differenze sono ottenute utilizzando il modello con interazioni fra variabili.

	Diff.za v. Italiani	
	Prima gen.	Seconda gen.
Nord Ovest	-8,5	-8,2
Nord Est	-7,2	-7,0
Centro	-4,1	-5,1
Sud	-5,1	-8,3
Sud e Isole	-0,7	-7,5
ITALIA	-5,3	-7,3

Tabella 2 – Differenze attese nei punteggi della prova di italiano fra studenti con background migratorio e studenti nativi (fonte: elaborazione su dati INVALSI).

I risultati sono indubbiamente interessanti; comunque, devono essere letti con prudenza, collocandone l'interpretazione all'interno di un più ampio contesto che tenga conto di tutte le informazioni rilevanti attinenti al più ampio contesto socio-economico. Si deve considerare, ad esempio, che gli studenti con background migratorio sono assai più numerosi, sia in termini assoluti che in termini relativi, nel Nord Ovest e nel Nord Est. In tali aree, i divari rispetto agli studenti nativi sono comunque degni di nota – nell'ordine di 7-8 punti – seppur generalmente inferiori a quanto ci si potrebbe attendere utilizzando una logica puramente sommativa degli effetti.

Nel Sud e nel Sud ed Isole, gli studenti con background migratorio sono meno numerosi, soprattutto quelli di prima generazione. Poiché i risultati sono rilevati al termine dell'ultimo anno della scuola secondaria di secondo grado, si può immaginare che tale gruppo abbia una natura più selezionata – sia costituito da studenti più dotati o motivati che hanno superato la selezione degli anni precedenti oppure che appartengono a gruppi nazionali europei con minori difficoltà di assimilazione della lingua italiana.

Del resto, al Sud il divario degli studenti di seconda generazione è simile a quello che si registra nelle regioni del Nord. Gli studenti di seconda generazione sono quelli che, in qualche modo, indicano la linea di tendenza del fenomeno (sono, ad esempio, il gruppo in crescita). Fra gli estremi del Nord e del Sud, il caso delle regioni del Centro presenta un certo interesse. Qui, effettivamente, i divari tendono ad essere più contenuti rispetto alla media nazionale, tanto per gli studenti di prima che di seconda generazione. Ancora una volta, per poter dare un senso compiuto a tale osservazione occorre approfondire il tema delle caratteristiche degli studenti – ad esempio, la loro composizione nazionale o il loro livello socio-economico (nota: il modello controlla per il livello socio-economico) – ma anche, eventualmente, quello delle caratteristiche dei sistemi sociali, istituzionali e/o scolastici insediati sul territorio.

Spunti di approfondimento

I risultati mostrano che il background migratorio è associato in genere ad un più basso livello di competenze nella lingua italiana, ma tale differenza **non è uniforme**: cambia, ad esempio, a seconda del territorio.[7]

Questo indica che le condizioni di fragilità non si sommano in modo meccanico. In alcuni contesti, gli svantaggi tendono a cumularsi; in altri, l'ambiente scolastico o territoriale parrebbero contribuire ad attenuarli.

Comprendere queste interazioni è essenziale per:

- individuare i contesti in cui le difficoltà si concentrano;
- progettare interventi mirati;
- evitare letture troppo semplicistiche dei fenomeni.

Come già sottolineato nel numero precedente, i modelli statistici non forniscono risposte definitive, ma permettono di **porre domande più precise** e di orientare l'analisi verso le situazioni di maggiore criticità.

È in questo spazio che può svilupparsi un dialogo proficuo tra analisi quantitativa e riflessione pedagogica, nella prospettiva di politiche educative più efficaci e inclusive.

Per chi è interessato ad approfondire il tema dei modelli di regressione con interazioni fra variabili:

- **Corbetta, P. (2014)**, Metodologia e tecniche della ricerca sociale, Il Mulino.
- **Barone, C. e Schizzerotto, A. (2016)**, Sociologia dell'istruzione, Il Mulino.

[7] Esistono interazioni significative anche fra background migratorio e tipo di scuola, e fra background migratorio e genere. Tali aspetti non sono approfonditi in questa sede per motivi di spazio. Il tema sarà comunque ripreso e sviluppato in edizioni successive della newsletter.

Appendice: tabella completa dei coefficienti del modello di regressione con interazioni

Prima generazione	-4,1	1,1	*
Seconda generazione	-5,1	0,6	*
Area geografica (rif. Centro)			
Nord Ovest	13,7	0,2	*
Nord Est	14,9	0,2	*
Sud	-2,6	0,2	*
Sud e Isole	-5,3	0,2	*
Genere (rif. Maschi)			
Femmine	7,6	0,1	*
Regolarità degli studi (rif. Regolare)			
In ritardo di almeno un anno	-9,9	0,2	*
Percorso di studi (rif. Istituto Tecnico)			
Liceo Tradizionale	21,7	0,2	*
Altro Liceo	3,2	0,2	*
Istituto Professionale	-15,3	0,2	*
Escs studente	1,3	0,1	*
Escs classe	7,4	0,2	*
*Interazioni (rif. Nativi * Centro)*			
Prima generazione * Nord Ovest	-4,4	1,0	*
Seconda generazione * Nord Ovest	-3,1	0,6	*
Prima generazione * Nord Ovest	-3,1	1,0	*
Seconda generazione * Nord Ovest	-1,9	0,6	*
Prima generazione * Nord Ovest	-1,0	1,2	
Seconda generazione * Nord Ovest	-3,3	0,8	*
Prima generazione * Nord Ovest	3,4	1,5	*
Seconda generazione * Nord Ovest	-2,5	0,9	*
Prima generazione * Femmine	-6,2	0,7	*
Seconda generazione * Femmine	-3,0	0,4	*
Prima generazione * Liceo Tradizionale	-3,6	1,1	*
Seconda generazione * Liceo Tradizionale	-2,6	0,7	*
Prima generazione * Altro Liceo	3,0	1,2	*
Seconda generazione * Altro Liceo	-0,5	0,8	
Prima generazione * Istituto Professionale	-1,5	0,9	
Seconda generazione * Istituto Professionale	-1,4	0,7	*
Prima generazione * escs studente	-0,6	0,4	
Seconda generazione *escs studente	-1,9	0,3	*
Prima generazione * escs classe	-0,7	1,1	
Seconda generazione *escs classe	0,5	0,7	
n. casi	412.620		
R2	26,4		
* effetti significativi al 95%			

A questo link è possibile recuperare i precedenti numeri della newsletter.