Quality & Quantity
https://doi.org/10.1007/s11135-026-02822-x

Investigating the effect of teacher practices on student achievement: insights from INVALSI data

Iacopo Moreschini[1] · Nunzia Brancaccio[2] · Maria Prosperina Vitale[3] · Isabella Sulis[4]

Received: 23 December 2025 / Accepted: 1 May 2026

Abstract
The present contribution aims at evaluating the effect of teaching practices on high school students' performance focusing on INVALSI mathematics and reading tests at grade 13. The student-level data refers to the 2021-2022 cohort and are used jointly with classroom-level data gathered by teachers' questionnaires. This allows us to consider the effect of instructional practices and teachers' qualification and experience over students' sociodemographic characteristics. Multilevel logistic models are adopted to assess factors influencing the probability of over- and under-achieving. These latter are defined accordingly to the INVALSI proficiency classification, considering test scores grouped in five classes. Main findings suggest that educational practices under investigation have not an effect on over-achieving, whereas some are associated with under-achieving. Furthermore, teachers' qualification and experience appear as key determinants for enhancing students' performance and preventing failures.

Keywords Teaching practices · Multilevel logistic models · INVALSI data · High school · Student achievement

✉ Iacopo Moreschini
i.moreschini@campus.unimib.com

Nunzia Brancaccio
nunzia.brancaccio@unina.it

Maria Prosperina Vitale
mvitale@unisa.it

Isabella Sulis
isabella.sulis@unica.it

1 Department of Sociology and Social Research, University of Milan-Bicocca, Piazza dell'Ateneo Nuovo, 1, Milan 20126, Italy

2 Department of Social Sciences, University of Naples Federico II, Vico Monte della Pietà, 1, Naples 80138, Italy

3 Department of Political and Social Studies, University of Salerno, Via Giovanni Paolo II, 132, Fisciano 84084, Italy

4 Department of Political and Social Sciences, University of Cagliari, Via Sant Ignazio da Laconi, Cagliari 09123, Italy

Published online: 20 May 2026

1 Introduction

Teachers are one of the most important resource that the educational system provides to its students. Scholars have highlighted that, beyond their behaviour in the classroom, their subject knowledge, experience, level of education, and professional training are factors capable of influencing students' outcomes (Burroughs et al. 2019). With regard to behaviour in the classroom, teachers' ability to adapt suitable practices for a specific classroom context is recognized as a key aspect for improving students' engagement and, therefore, their academic outcomes (Tomaszewski et al. 2022). Innovative teaching practices can serve as a strategic tool for improving educational outcomes in any subjects (Chinelo 2020), increasing the rate of students achieving above-average results, and simultaneously reducing the likelihood of under-achieving.

Nonetheless, there are well documented disparities observed in the Italian education system related to students' family socioeconomic conditions (Contini et al. 2018; Pensiero et al. 2019; Sulis et al. 2020), gender (Contini et al. 2017; Tocchioni et al. 2025) and the presence of a migratory background on students' educational choices and, indeed, on their educational success (Barone 2009; Tocchioni et al. 2025). Moreover a high variability is observed at the class and the school level in students' results (Invalsi 2019, 2024) that exacerbates disparities. Within this framework, teachers' practices, experience, and qualifications represent strategic assets for policymakers seeking to promote policies aimed at reducing structural inequalities.

The spread of innovative instructional practices is naturally influenced by the distribution of the school's workforce according to some teachers' characteristics (e.g., age, subject), including students attending a certain track, the types and number of tracks at schools. These factors permit to classify high schools and to shape the attractiveness of certain schools over others for students with similar characteristics. Keeping all this in mind, transformations in the classroom environment have been also observed, driven by the integration of ICT, which was significantly accelerated by the COVID-19 pandemic period.

Building on these considerations, the paper investigates the influence of innovative teaching practices on students' academic outcomes, operationalised here as over-achievement or under-achievement based on the competence-level classification developed by the Italian National Institute for the Evaluation of the Education and Training System (INVALSI).[1] First, the analysis focuses on assessing students' performance, shifting the emphasis from modelling average competences to identifying factors that have the greatest impact on over- and under-achievement. Second, focusing on the analysis on teachers' characteristics in terms of experience, the qualification and the effectiveness of new instructional practices permit to shed light on the role of these strategic assets in promoting excellence and preventing poor performances in Italian high schools. These aspects have received considerable

[1] INVALSI is the public body responsible for the external standardised assessment of student learning outcomes across the Italian school system. Its large-scale assessments are census-based and are administered annually at key educational stages (grades 2, 5, 8, 10, and 13). Tests are administered directly by schools and, from Grade 8 onward, are computer-based. In addition to the census administration, each year a subset of high schools, including grades 10 and 13, is selected for the national sample; in these schools, computer-based tests are administered under the supervision of external inspectors in order to reduce cheating and ensure more controlled administration conditions. The assessments measure competencies in reading, mathematics, and english, and scores are estimated using Item Response Theory (IRT). Specifically, Rasch-family models (Invalsi 2016) are adopted and scores are classified in five competence-level categories.

attention in the recent literature (Sulis et al. 2020; Porcu et al. 2023), as policy efforts have mainly been tailored to optimising average competencies within the student population. Third, the analysis of factors influencing poor versus high-quality competences has been recently challenged by new regulations of the Italian Ministry of Education on educational standards.[2] The debate, emphasising the need for personalising instruction in schools, shifts the attention to the definition of educational programs addressed to foster excellence and reduce disparities.

Building on this framework, the main research questions addressed in the present contribution are:

1 To what extent do teachers' characteristics, in terms of professional experience and training, are associated with students' achievement, and how do these effects vary between the two core disciplinary areas of reading and mathematics?
2 To what extent shared instructional practices and teaching strategies are associated with patterns of student success and failure?
3 To what extent ICT tools may play a key function as facilitators or barriers to students' achievement within contemporary high school learning environments?

To address these research questions, the analysis draws on INVALSI data from the 2021–2022 cohort of 13th-grade students, combined with classroom-level teacher survey data, allowing the matching of student outcomes with instructional practices and teacher characteristics. The rest of this paper is organised as follows. Section 2 is devoted to providing a brief literature review. Section 3 presents the data and the methodological approaches adopted in the analysis. Results from descriptive analysis and modelling approaches are presented in detail in Sect. 4. In Sect. 5, main findings and concluding remarks are discussed.

2 Literature review

Since the Coleman Report (Coleman et al. 1966), the impact of teachers on students' academic achievement has been the focus of research that seeks to highlight how the learning environment and the educational relationships between students and teachers are the elements capable of limiting the impact of extracurricular factors on achievement. Teacher effectiveness or quality, as the ability of teachers to influence student outcomes, is therefore a widely explored topic. This concept is subject to a multiplicity of definitions, as summarised by Burroughs et al. (2019). On the one hand, some teacher characteristics, which include experience (years in the teaching profession), professional knowledge (educational qualification, participation in further training) and content knowledge, have been found to be positively associated with students' achievement. On the other hand, teachers' behaviours, which form a bridge to another strand of the literature that investigates teaching practices, often expressed through the lens of the relationship between teachers and students and the support they provide them, can influence academic results and student involvement in school activities (Tomaszewski et al. 2022).

[2] For details see: https://www.mim.gov.it/-/scuola-approvato-al-senato-il-ddl-per-il-riconoscimento-degli-studenti-ad-alto-potenziale-cognitivo-valditara-la-personalizzazione-della-didattica-e-l.

However, the support given by teachers to students is only the surface of the structured interactions that take place in the classroom environment. Teaching work takes shape in specific instructional practices, the impact of which on achievement is, however, analysed by scholars in a fragmented manner, consistently with the didactic specificities of individual subjects. Yu and Singh (2018), for example, compare two different methods of teaching mathematics, observing how teaching focused on procedures determines negative effects on learning, while concept-based explanations have a positive effect.

Beyond practices that are specific to a single subject, there is growing interest in instructional practices that can, instead, be transversal across several subjects. Key examples include group and cooperative work practices, flipped classroom methodologies, and inquiry-based or laboratory-oriented approaches. Bilici and Yilmaz (2024), using a quasi-experimental design, focus on collaborative storytelling, a methodology through which students are asked to produce, in groups, outputs to be presented to the whole class, finding a positive effect on achievement in biology. Abukari et al. (2023) assessed the influence of cooperative learning methodology on chemistry achievement, finding a significant increase in achievement in their treated groups. These results are consistent with meta-analytical evidence that cooperative and collaborative learning are on average associated with small-to-moderate gains in academic achievement across subjects and levels of schooling (Kyndt et al. 2013; Yaşar et al. 2024).

Moreover, a recent meta-analysis synthesises a large body of evidence on the effectiveness of the flipped classroom methodology in various disciplines (Hew et al. 2021), in line with the broader literature reporting generally positive, though heterogeneous, effects of flipped learning on students' achievement and motivation (Strelan et al. 2020; Zheng et al. 2020). Alongside cooperative and flipped approaches, inquiry-based and laboratory-oriented instructional practices represent another family of cross-curricular methodologies. Several studies indicate beneficial effects of these practices on conceptual understanding, higher-order thinking, and, under well-structured conditions, on achievement outcomes in science and related domains (Kaçar et al. 2021; Villanea 2023; Arifin et al. 2025). Nonetheless, some large-scale studies warn that poorly implemented inquiry may fail to translate into higher performance (Jerrim et al. 2022).

What emerges from the literature is that a solid body of research on teacher characteristics is accompanied by a more fragmented focus on what actually happens inside classrooms and on the impact that teaching practices have on student outcomes. Studies often consider a single subject and therefore are limited to compare the effectiveness of teaching methodologies across different subjects. This gap appears even more pronounced in the context of the Italian education system, where systematic, cross-subject evidence on instructional practices and their effects on achievement is particularly scarce. Within this theoretical framework, the main purpose of the present study is to integrate information from the INVALSI teachers' survey with data on students' competences in order to shed light on how individual characteristics, instructional practices, and teachers' experiences interact in shaping educational success.

3 Data and methods

3.1 Data

INVALSI sample data on 13th-grade students from the 2021–2022 student cohort were gathered and matched with teacher-level data collected at the classroom level. The sample consists of 11,196 students, grouped into 579 classes across 288 schools. The teacher survey collects information on instructional practices and individual characteristics from 920 teachers (464 literature teachers and 456 mathematics teachers) in 505 classes across 260 schools, allowing student records to be matched with subject-specific teacher information.

Missing data at the student level arise from two main sources. First approximately 10.5% of students have missing values on the Index of Economic, Social and Cultural Status (ESCS index), which summarises socioeconomic status based on parental occupation and education, and household ownership. These missing values were addressed through hierarchical imputation, replacing them with the mean computed for students sharing the same classroom and same migratory background (considering the three categories native, first-generation, or second-generation). The imputation procedure adopted permitted a complete recover of values affected by missingness in the ESCS index. Missing observations affect the reading test score (6.7%), the mathematics test score (6.9%), and the migration background control variable (1.2%), as shown in Table 10. The other factor of the presence of missing observations is related to the matching procedure between students- and teacher-level data. This matching results in the loss of 115 classes for literature teachers and 123 classes for mathematics teachers. This source of missingness was considered structural, and the reference samples for the analysis include only students for whom the teacher survey was also completed (n= 8960 for the reading sample and n = 8784 for the mathematics sample). The combination of structural missing data due to unobserved teachers and the list-wise deletion of missing observations in tests scores and migration background results in two final analytical samples with numerosity of 8320 for the reading sample and 8,171 for mathematics sample, with a missingness rate of respectively 8–7% with respect to the reference datasets. Missing data are not completely randomly distributed across the sample: they are more concentrated among students enrolled in vocational and technical tracks and in southern regions (Table 11). Nonetheless no systematic differences emerge in socioeconomic background measured by the ESCS index between students with missing observations, even controlling for the school's track.

3.2 Measures

The INVALSI competence assessment programme annually covers the entire population of Italian students in grades 5, 8, 10, and 13. Similar to the OECD PISA framework, where test scores are mapped onto competence levels, INVALSI scores can likewise be classified into five competence levels, each associated with qualitative descriptors of students' proficiency.[3] For mathematics and reading comprehension at the 13th grade, scores that do not reach the 3rd level are considered below the minimum competence level that is expected from students at the end of the high school (Invalsi 2024). These students are referred to

[3] For details: https://invalsi-areaprove.cineca.it/index.php?get=static&pag=g13_descrittori_qualitatitvi.

as under-achievers. On the other side, students who reach the 5th level are considered as over-achievers.

Furthermore, the INVALSI teacher survey includes a section designed to collect information on reading and mathematics instructional practices. The following analysis focuses on items from the teaching questionnaire, which use Likert scales to measure the frequency with which teachers implement specific behaviours and instructional strategies. Specifically:

- Ten items measure the frequency of ICT usage by teachers, covering the use of personal and school computers, interactive whiteboards (IWB), personal and school tablets, smartphones, personal and institutional e-learning platforms, and personal or institutional educational software. The original three-point scale (*regular use*, *occasional use*, *no use*) was dichotomised by coding *regular use* as 1 and all other categories as 0. The choice to dichotomise responses has been made by the aim of capturing the association of a regular use of ICT on educational success, avoiding to take into consideration any form of sporadic use. This option clearly disentangles didactic approaches that systematically rely on the use of ICT from the other. An Item Response Theory model (IRT) (Birnbaum 1969) was then estimated to summarise the ten indicators into a single latent measure at teacher-level of attitude to ICT usage. The use of such an approach permitted to jointly assess the combined and systematic use of ICT to shape different teachers' profiles.
  The main advantage of IRT is its capability to provide a unique score of the latent trait for those teachers who exhibit exactly the same response pattern, reflecting the uniqueness in the attitude to the use of ICT for teaching related to each single pattern of the observed responses. Thus, the use of IRT for dichotomous items represents an advantage over the Classical Test Theory approach, which assigns the same score to items that have different rates of positive answers and treats items as if they were numeric. Moreover, this approach identifies informative and non-informative items, weighting them according to their level of informativeness.
- Four items measured on a four-points Likert scale (*never*, *sometimes*, *often*, *always* or *almost always*) related to the frequency of teaching practices different from traditional lecturing. These include cooperative activities among peers, classroom presentations of multimedia products, flipped classroom or flipped teaching methodologies, and laboratory-based activities.
- Two items measured on a four-points Likert scale related to the frequency of assessment methods different from individual evaluations, that is, group works evaluation and the use of ICT platform for testing and evaluation.

As control variables, additional information on teachers' sociodemographic characteristics (gender and age), educational qualifications, years of service, and contract duration helps to shed light on multiple sources of variability that may be associated with students' success in high schools.

3.3 Methods

A multilevel logistic regression (Goldstein 2011) is used to model the probability of over- or under-achieving in mathematics and reading (Sect. 3.2). The multilevel logistic regression approach makes it possible to effectively deal with the hierarchical structure of the INVALSI data, where students (level one units) are grouped into classes (level two units) that share the same teachers and classroom environment. The model can be specified as follows:

$$\mathrm{logit}(P(y_{ic}=1)) = \beta_0 + x'_{ic}\beta_1 + x'_c\beta_2 + u_c \tag{1}$$

In 1, $P(y_{ic}=1)$ represents the probability that a student *i* in class *c* is an over- or underachiever. The term x'_{ic} is the vector of individual characteristics of student *i* in class *c*, namely socioeconomic status (SES), gender, macro-regional area of residence and migration background, while x'_c represents the vector of variables related to teaching practices and teacher characteristics observed for class *c*. Additionally, u_c represents the normally distributed random intercept term that captures the association between students belonging to the same class. It permits to take into account unobservable factors explaining divergences in the results at the class level. The posterior estimations of the level-2 random intercepts (empirical Bayes estimates) are considered as adjusted indicators of the so called *class effect* in explaining divergences in students' competences (Goldstein and Spiegelhalter 1996; Leckie and Goldstein 2019) and they are used in the literature to assess divergences between classes.

In addition, to derive latent trait scores representing teachers' attitude to ICT use, it was applied a dichotomous IRT model to the ten dichotomised indicators which capture teachers' attitude to regularly integrate digital tools into their instructional routine. Given the binary nature of the recoded items, the analysis relies on the Two-Parameter Logistic Model–2PL IRT (Birnbaum 1969). In IRT models, the probability of endorsing an item depends on both respondent characteristics (person parameters) and item-specific features (item parameters). Under the 2PL, the Item Response Function (IRF) for item *i* takes the form:

$$P(Y_{ti}=1 \mid \theta_t) = \frac{\exp[\alpha_i(\theta_t-\delta_i)]}{1+\exp[\alpha_i(\theta_t-\delta_i)]}, \tag{2}$$

where the person parameter θ_t denotes the normally distributed individual latent trait level, namely the teacher's attitude to use ICT. α_i is the item discrimination parameter, and δ_i is the item location parameter. The discrimination parameter α_i governs the steepness of the IRF around the location parameter. Higher values of α_i correspond to steeper curves, meaning that the item more effectively differentiates between teachers whose latent ICT usage is near δ_i. Conversely, low-discrimination items produce flatter curves, providing weaker information about differences in the underlying trait. In practice, discrimination parameters typically assume moderate positive values (Hambleton et al. 1991).

The item location parameter δ_i retains the same interpretation as in the Rasch (1PL) model (Rasch 1980), representing the location on the latent continuum, where the probability of provide a positive answer to an item reaches.5. A feature of the IRT models is that person location parameters and item location parameters are expressed in the same metric and this permits to interpret the latter as the minimum level of latent trait required to a respondent to have a probability greater that.5 to endorse it. For this reason, the item location parameter is also known as item difficulty parameter. The term, borrowed from the context of standard test analysis (De Ayala 2013), where these models were originally developed, in this context indicates that only teachers whose ICT attitude levels exceed the item's location value are likely to respond 1. Therefore, the higher the item parameter for any item i, the more difficult it is for a teacher to reach the minimum level of latent trait required to answer it positively. Conversely an item with a low location parameter (also called easy item) reflects an activity in the use of ICT that even teachers with lower ICT attitudes often endorse. Unlike in the 1PL model, however, IRFs in the 2PL may intersect, implying that the relative "difficulty" of items, in terms of probability to provide a positive answer, can vary across different regions of the latent trait (Baker and Kim 2004). This property reflects the greater flexibility of the 2PL in accommodating heterogeneous item behaviour.

Model estimation yields item parameters and empirical Bayes estimates of θ_t are predicted at posterior on the basis of the prior distribution of θ_t. These latter represent continuous standardized scores capturing teachers' propensity to integrate ICT in their professional practice. The latent trait scores are computed moving from the teachers' questionnaire and are incorporated as class level covariates in the models. The estimated latent trait score is used as predictor of teachers' attitude to ICT use in the multilevel regression model in Eq. 1.

4 Results

4.1 Insights from descriptive statistics

The distribution of students' academic performance highlights inequalities related to gender, migratory background, geographic location, and socioeconomic status. Descriptive statistics (Table 1) presented separately for the reading and mathematics samples, shade light on how disparities are deeply rooted in the Italian educational system. These aspects are reflected not only in students' competences but also in the teaching practices and assessment methods

Table 1 Students' characteristics for reading (n= 8320) and mathematics (n= 8171)

	Reading	Mathematics
Gender	%	%
Male	47.1	47.7
Female	52.9	52.3
Migratory background	%	%
Native	91.8	91.5
1st generation	2.7	2.7
2nd generation	5.5	5.8
Region	%	%
North	41.7	41.1
Centre	19.2	21.3
South and Islands	39.1	37.6
Track	%	%
Scientific track	24.6	24.3
Classic and Linguistic track	14.2	14.6
Other general track	14.7	13.5
Technical	31.7	32.9
Vocational	14.8	14.7
Achievement	%	%
Low	41.2	47.7
High	8.1	15.8

employed across different school tracks. Gender differences emerge clearly in both subjects, although with distinct patterns. In reading, males are over-represented in the under-achiever category (47.4%), while 9.6% of females are classified as over-achiever. In mathematics, the pattern is reversed: among over-achievers, males are over-represented in mathematics with 20.8% of males classified in the top categories versus 11.4% of females. The gender gap in reading appears narrower when compared to the gender gap in mathematics (Table 2).

Migratory background emerges as a critical factor in both low and high performance. First-generation students, born abroad to foreign-born parents, face the steepest challenges, with half classified as under-achievers in reading (50.0%) and an even higher share in mathematics (58.9%). High achievement among this group is rare, with just 0.8% excelling in reading and 11.0% in mathematics. Second-generation students, born in Italy to foreign-born parents, show some improvement but remain behind their native peers. For this group, 51.7% are under-achievers in reading and 49.9% in mathematics, while only 4.4% and 12.8% are over-achievers in these subjects, respectively. In contrast, native students - those born in Italy with at least one Italian-born parent or a mixed background - perform substantially better, with smaller proportions of under-achievers (40.4% in reading and 47.3% in mathematics) and higher shares of over-achievers (8.5% in reading and 16.2% in mathematics).

Geographic differences further exacerbate these inequalities: students from the South and the Islands are overrepresented among under-achievers, with 54.2% of students from this area falling into this category in reading and 65.1% in mathematics, while in the North students perform better, with under-achiever rates of 29.8% and 33.1%, respectively in read-

Table 2 Under-achievement and over-achievement rates for reading (n= 8320) and mathematics (n= 8171)

	Low Reading	High Reading	Low Mathematics	High Mathematics
Gender	%	%	%	%
Male	47.4	6.4	43.2	20.8
Female	35.8	9.6	51.8	11.4
Migratory background	%	%	%	%
Native	40.4	8.5	47.3	16.2
1st Generation	50.0	0.8	58.9	11.0
2nd Generation	51.7	4.4	49.9	12.8
ESCS quartiles	%	%	%	%
Q1	59.3	2.3	62.2	8.0
Q2	46.5	5.3	52.3	12.0
Q3	36.9	8.3	43.7	15.5
Q4	22.3	16.4	31.9	28.1
Geographical macro-areas	%	%	%	%
North	29.8	10.8	33.1	23.7
Centre	40.1	8.0	45.4	16.0
South and Islands	54.2	5.3	65.1	7.2
Total	41.2	8.1	47.7	15.8

ing and mathematics. Conversely, the North hosts a larger proportion of over-achievers, particularly in mathematics (23.7%), compared to the South, where the percentage drops to just 7.2%.

The role of socioeconomic status, as captured by the ESCS index, provides evidence of a strong mechanisms of intergenerational transmission of inequalities in education (Barone 2009; Fabrizi et al. 2024; Lo Cicero 2024). In reading, the students in the bottom quartile of the socioeconomic distribution show a proportion of under-achievers of 59.3% compared to the 22.3% among those in the top quartile. Conversely, high achievement is concentrated among the most advantaged students, with 16.4% reaching top performance, compared with only 2.3% among the most disadvantaged. A similar pattern emerges in mathematics, where students in the bottom quartile display higher shares of low achievement reaching 62.2%, compared to the 31.9% rate among those in the top quartile. At the same time, high achievement is concentrated among the most advantaged students (28.1%), while remaining comparatively rare among those in the lowest quartile (8.0%) (Table 2).

These differences are not bounded to academic outcomes in terms of competences, but they are deeply reproduced in the educational choices of track. Evidence from sociological research shows that students from socioeconomically advantaged families are more likely to enrol in prestigious general track (lyceum) schools (Pensiero et al. 2019). For instance, scientific tracks, where the average ESCS is about 0.5, and classical and linguistic tracks, with an average ESCS of 0.4, attract students from more privileged backgrounds and boast the highest shares of over-achievers, particularly in mathematics (38.3% in scientific tracks) and reading (15.6% in classical and linguistic tracks) (Table 3). In contrast, technical and vocational schools, which primarily attract students from disadvantaged background (average ESCS of −0.3 and −0.6, respectively), show notably high proportions of under-achievers, reaching the value of 79.0% of students classified as under-achievers in reading and 83.8% in mathematics in vocational schools.

Table 4 reports the frequency with which teachers employ each teaching practice across school tracks, highlighting some disparities. Technical and vocational schools, aligned with their practical orientation, place greater emphasis on collaborative and hands-on learning methods. Peer activities, for instance, are a central practice in vocational schools, with

Table 3 Students characteristics for reading (n= 8320) and mathematics (n= 8171) by track

	Scientific track		Classic/Linguistic track		Other general track		Technical		Vocational	
	Read	Math	Read	Math	Read	Math	Read	Math	Read	Math
Gender (%)										
Male	53.8	54.3	23.6	23.2	16.3	17.8	60.8	60.7	59.6	59.2
Female	46.2	45.7	76.4	76.8	83.7	82.2	39.2	39.3	40.4	40.8
Migratory background (%)										
Native	95.5	95.1	94.5	94.4	92.9	93.1	88.6	88.0	88.7	89.5
1st Generation	1.2	1.2	1.7	1.8	2.1	1.7	4.1	4.3	3.7	3.3
2nd Generation	3.3	3.7	3.8	3.9	4.9	5.2	7.3	7.7	7.6	7.2
Achievement (%)										
Low	17.8	17.1	19.0	49.6	42.3	62.2	51.3	47.4	79.0	83.8
High	15.6	38.3	16.6	9.8	6.2	5.9	3.0	12.9	0.3	0.6
Students' SES (μ)										
ESCS	0.5	0.5	0.4	0.4	−0.03	−0.02	−0.3	−0.2	−0.7	−0.6

Table 4 Use of practices by reading teachers (n= 464) and mathematics teachers (n= 456)

	Scientific track		Classic/ Linguistic track		Other general track		Technical		Vocational		Total	
	Read	Math	Read	Math	Read	Math	Read	Math	Read	Math	Read	Math
Peer activities (%)												
Never or almost never	13.8	27.8	17.0	23.7	15.8	20.2	10.1	21.6	3.8	15.7	11.6	22.1
Sometimes	50.1	43.2	43.1	39.7	54.4	50.2	41.2	42.5	34.9	34.2	44.4	41.9
Often	26.7	19.7	20.8	32.8	19.8	24.7	36.6	30.5	53.2	39.8	32.6	29.2
Always or almost always	9.4	9.3	19.1	3.8	10.0	4.9	12.2	5.4	8.1	10.3	11.4	6.9
Classroom presentations (%)												
Never or almost never	16.0	42.1	15.4	27.3	18.2	31.0	3.8	43.8	8.0	34.5	11.1	37.6
Sometimes	43.9	40.4	45.9	54.1	39.1	55.6	48.1	44.5	29.0	46.0	42.1	46.7
Often	32.8	17.5	25.1	15.5	34.8	13.4	39.1	6.7	45.5	19.5	36.2	13.7
Always or almost always	7.3	0.0	13.6	3.0	7.9	0.0	9.0	5.0	17.5	0.0	10.5	1.9
Flipped classroom (%)												
Never or almost never	37.8	58.2	37.4	46.9	30.8	53.7	23.0	41.9	30.5	36.6	30.9	47.1
Sometimes	43.1	27.6	41.9	45.8	44.2	42.5	54.1	47.8	44.7	43.1	46.8	41.3
Often	17.2	14.2	17.3	7.3	23.0	3.7	18.7	8.1	22.8	18.1	19.5	10.6
Always or almost always	1.9	0.0	3.4	0.0	2.0	0.0	4.1	2.1	2.0	2.2	2.8	1.1
Laboratorial activities (%)												
Never or almost never	33.1	25.3	30.0	41.3	29.2	41.1	20.5	53.3	23.9	30.3	26.6	39.2
Sometimes	42.1	50.0	42.3	45.6	44.1	47.8	51.9	39.5	54.5	51.1	47.6	46.1
Often	20.0	24.4	24.5	12.5	22.9	8.9	24.4	7.3	18.5	15.0	22.1	13.6
Always or almost always	4.9	0.3	3.2	0.5	3.9	2.2	3.2	0.0	3.0	3.6	3.7	1.1
Type of class (%)												
Mostly in person	74.8	75.3	73.2	79.8	75.2	79.4	74.9	70.9	84.0	62.9	76.3	72.9
Mostly remote	0.0	0.6	0.0	0.0	1.0	0.0	0.5	1.1	2.3	0.0	0.7	0.5
Equally in person and remote	3.1	1.9	0.0	1.6	1.6	2.2	1.7	1.0	2.9	5.4	2.0	2.3
Only in person	22.1	22.2	26.8	18.6	22.2	18.5	22.9	27.1	10.8	31.7	21.0	24.4

53.2% of reading teachers and 39.8% of mathematics teachers report using them often, with an additional 8.1% and 10.3% using them always. By contrast, these activities are far less common in general tracks, where collaborative methods are not as prioritised. A notable exception is mathematics in the classic and linguistic tracks, where 36.6% of teachers report using them at least often. Laboratory activities appear to be generally not very widespread, though they are more common among reading teachers. They are used at least often by about 25% of teachers in the general tracks and in the technical track, whereas they seem to be less explored in the vocational track (21.5%). They are clearly less common in mathematics teaching: only in the scientific track they reach a level of use comparable to that observed in reading instruction. Another exception is the vocational track, where they are used at least often by 18.6% of mathematics teachers. The same difference in use across subjects can be observed for students' presentations of multimedia work ("classroom presentations"). This practice is extremely widespread among reading teachers, especially in the vocational track, where it is used at least often by 63.0% of teachers, and in any case by more than 30% in all other tracks as well. Mathematics teachers, by contrast, appear to be less inclined to use this teaching method. The flipped classroom is scarcely adopted across all tracks. Even in vocational schools, which record the highest usage, only 24.8% of reading teachers report using it regularly, while its adoption in general track schools is almost absent (Table 4). This table also reports the predominant mode of instruction, distinguishing among different degrees of remote teaching and thereby capturing a data collection period still shaped by the effects of Covid-19 restrictions. The fact that the emergency phase of the pandemic had largely passed is reflected in the very high proportion of teachers reporting that classes were conducted either exclusively or predominantly in person, ranging from 94% to 98.4% across all tracks and reaching 100% among reading teachers in classic and linguistic tracks. At the same time, the technical and vocational tracks appear to be those in which distance teaching remained comparatively more prevalent during the period under examination.

Assessment methods display a less clear pattern across tracks, but marked differences still emerge between reading and mathematics teachers. Group evaluations appear largely an occasional rather than systematic assessment, most frequently employed among reading teachers in scientific track and in vocational schools (61.2% and 69.6%, respectively). By contrast, mathematics teachers are substantially less likely to rely on group evaluation: in the scientific and technical tracks, 59.1% and 56.7% of teachers, respectively, report never or almost never using this strategy. Only in the other general tracks, which are less strongly oriented towards mathematics, around 60% of teachers report using group evaluation at least occasionally. The use of digital platforms for assessment, such as Kahoot or Moodle, is even less widespread across all tracks(Table 5). The figures reported in Tables 4 and 5 do not represent the proportion of students exposed to these practices, as student exposure also depends on how students are distributed across tracks. The corresponding student-level figures are reported in Tables 12 and 13.

4.2 Operationalising ICT usage latent trait

Table 6 reports the proportion of teachers who declare regular use of each ICT tools in reading and mathematics, respectively. There is a substantial variation across items: some tools, such as personal and school computers or interactive whiteboards, are regularly used by a

Table 5 Use of evaluation strategies by reading teachers (n= 464) and mathematics teachers (n= 456)

	Scientific track		Classic/ Linguistic track		Other general track		Technical		Vocational		Total	
	Read	Math	Read	Math	Read	Math	Read	Math	Read	Math	Read	Math
Group evaluations (%)												
Never or almost never	26.5	59.1	29.3	44.6	32.2	34.1	19.7	56.7	17.0	43.3	24.0	49.9
Sometimes	61.2	36.4	49.1	47.7	48.2	59.9	56.7	36.6	69.6	40.8	57.7	42.2
Often	10.5	4.5	19.5	7.7	17.4	6.0	20.7	5.9	13.4	16.0	16.4	7.7
Always or almost always	1.9	0.0	2.1	0.0	2.2	0.0	2.9	0.8	0.0	0.0	1.9	0.3
ICT for evaluation (%)												
Never or almost never	58.6	63.0	52.9	56.1	57.6	68.3	48.8	55.6	43.0	48.1	52.0	57.8
Sometimes	33.3	30.6	33.5	31.8	28.1	23.7	31.4	35.0	37.4	29.6	32.7	30.9
Often	6.2	6.5	12.1	12.1	12.8	8.0	16.6	5.5	19.6	21.1	13.5	9.8
Always or almost always	1.9	0.0	1.6	0.0	1.4	0.0	3.3	4.0	0.0	1.3	1.9	1.4

Table 6 Proportion of ICT usage across reading and mathematics teachers

ICT Tool	Reading	Mathematics
Personal computer	0.66	0.56
School computer	0.66	0.55
Interactive whiteboard	0.60	0.66
Personal tablet	0.31	0.34
School tablet	0.08	0.06
Smartphone	0.37	0.29
Personal e-learning platforms	0.28	0.22
Institutional e-learning platforms	0.40	0.29
Personal educational software/applications	0.32	0.38
Institutional educational software/applications	0.31	0.32
Cronbach's α	0.65	0.62

Table 7 IRT 2PL parameter estimates for ICT usage items

	Reading		Mathematics	
ICT tool	δ_i	α_i	δ_i	α_i
Personal computer	−0.928	0.814	−0.353	0.826
School computer	−5.702	0.118	−0.596	0.317
Interactive whiteboard	−0.701	0.609	−1.312	0.538
Personal tablet	1.322	0.652	1.347	0.524
School tablet	3.804	0.704	4.577	0.643
Smartphone	0.775	0.787	1.420	0.690
Personal e-learning platforms	0.698	2.482	1.102	1.735
Institutional e-learning platforms	0.362	1.688	0.793	1.727
Personal educational software/applications	0.527	2.946	0.408	1.952
Institutional educational software/applications	0.615	2.149	0.576	2.488

large percentage of teachers in both subjects, whereas others, including tablets and educational software, are adopted on a regular basis only by a small minority.

Internal consistency is acceptable (reading: α Cronbach = 0.65; mathematics: α Cronbach = 0.62), an expected result given the diversity of ICT resources considered. The application of an IRT 2PL model, which enables the definition of a latent construct that appropriately weights items according to their difficulty and discrimination parameters, allows a more precise investigation of measurement accuracy along the latent trait.

The 2PL was estimated separately for reading and mathematics teachers. Table 7 reports the estimated item (δ_i) parameters and discrimination (α_i) parameters for each ICT tool. The main results reveal marked heterogeneity in the item behaviour across the two groups of teachers. A subset of tools–particularly personal and institutional e-learning platforms, as well as educational software–exhibit high discrimination values, indicating that they are especially effective in distinguishing teachers along the central region of the latent ICT usage trait. Conversely, tools characterised by very low rates of regular use, such as school tablets, exhibit a high location parameter underlining the high difficulty to be score positively and contributing information primarily at the upper tail of the latent trait distribution. More widely used resources, including personal and school computers or interactive whiteboards, show moderate values of both item parameters, providing stable information across a broader range of ICT usage levels.

The Item Information Curves –IICs– highlight a consistent pattern across reading and mathematics teachers (Fig. 1). In both subjects, the most informative items are those related to e-learning platforms and educational software (I_7–I_8-I_9-I_{10}), which show pronounced information peaks around the central region of the latent trait. These items, which show high discrimination parameters, are therefore crucial for distinguishing teachers with a medium attitude of ICT usage. Items associated with rarely used tools, such as school tablets, exhibit high difficulty and contribute information only at the upper end of the trait distribution, while widely used tools (e.g., personal or school computers, interactive whiteboards) provide poor but relatively broad information. The Test Information Functions –TIFs– confirm that, for both groups, measurement precision is maximised around $\theta \in [0, 1]$, decreasing sharply in the tails. The mathematics scale yields a slightly lower overall information level

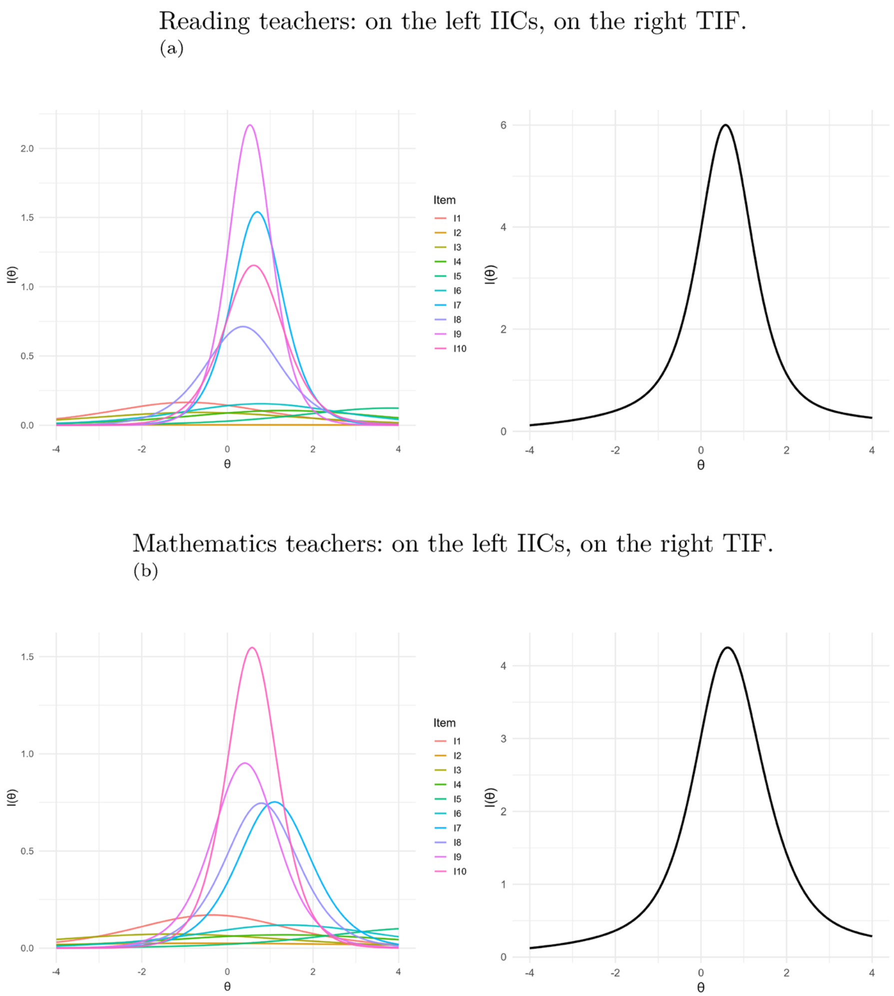

Fig. 1 IRT-based information functions for ICT usage items: **a** Reading, **b** Mathematics teachers

than reading, in line with its lower internal consistency. Overall, the ICT usage scale is most reliable for discriminating among teachers with average levels of the digital engagement.

4.3 Multilevel models results

Different sets of multilevel logistic models were conducted to evaluate how teacher characteristics and instructional practices are associated with the likelihood of under- or over-achievement in reading and mathematics. Table 8 presents the log-odds coefficients after controlling for student characteristics and academic tracks. Reported log-odds should be compared only across models referring to the same subject, since the analytical samples for reading and mathematics differ in size.

With regard to teacher characteristics, teacher experience, measured by the number of years in the teaching job, is positively associated with reading achievement. Each additional year of experience reduces the odds of under-achieving in reading by 1.13%, and increases the odds of over-achieving by the same amount. Conversely, teachers' gender and qualification influence the likelihood of over-achievement in mathematics. Having a female teacher increases the odds of over-achieving in maths by 34% with respect to have a male teacher, while having a teacher with education above a master's degree boosts the odds by 56%. The educational level of teachers is also associated with the odds of under-achievement in reading. Teachers with qualifications exceeding a master's degree reduce the odds by 26%. Lastly, the coefficients for employment conditions, operationalised by contract duration, suggest the importance of teacher continuity. Students taught by teachers with annual contracts have a 57% higher odds of under-achieving in reading and a 176% higher odds of under-achieving in mathematics compared to those taught by teachers with permanent contracts. The same contract duration is associated with a 59% decrease in the odds of over-achieving in mathematics. Similar associations with mathematics performance are shown by shorter contracts, which increase the odds of under-achieving by 158% and reduce the odds of over-achieving by 78%.

Although the full models consider various teaching and assessment practices, the estimated coefficients are significantly different from zero for only a few, and none of these show positive associations with over-achievement. Notably, flipped classroom methods, an instructional practice where students research topics independently before classroom discussions, are linked to poorer outcomes. Students who are often or always exposed to flipped classrooms see their odds of under-achieving in reading to increase by 35% compared to those who are never exposed to the practice. For mathematics, the negative association is clearer in the odds of over-achieving, which decreases by 33% for students exposed to such a practice even occasionally. Other assessment strategies, such as group work evaluations and ICT platform testing, are also negatively associated with maths achievement. Group work assessments increase the odds of under-achieving by 36% when used occasionally and by 52% when used often or always. Furthermore, even occasional reliance on ICT-supported testing increases the odds of under-achieving in maths by 31% compared to never using it.

The estimated coefficients for the Teacher ICT Use scale, which provides information on the attitude to the joint use of multiple technologies for teaching, have been found not to be significantly associated with any of the considered outcomes. This result was confirmed using alternative model specifications which substituted the single joint-use scale with the

Table 8 Estimated coefficients (log odds) for teacher characteristics and teaching practices on students' probability of under-achieving and over-achieving in reading (n= 8320) and mathematics (n= 8171)

	Reading		Mathematics	
	Under- achievement	Over- achievement	Under-achievement	Over-achievement
Intercept	−0.680**	−2.719***	−1.256***	−1.449***
Student characteristics				
Female	−0.187***	0.168*	0.581***	−0.780***
Migratory background ref.: Native				
1st Generation	0.367**	−1.728***	0.535***	−0.348
2nd Generation	0.519***	−0.476*	0.299**	−0.271
School track ref.: Other general track				
Scientific track	−1.621***	1.221***	−2.730***	2.551***
Classical/Linguistic track	−1.345***	1.067***	−0.985***	0.652***
Technical	0.349***	−0.652***	−0.575***	0.387
Vocational	1.606***	−2.813***	1.205***	−2.477***
ESCS	−0.171***	0.307***	−0.0461	0.174***
Region ref.: North				
South	1.490***	−1.076***	2.081***	−2.042***
Centre	0.757***	−0.557***	1.092***	−0.901***
Teacher characteristics				
Age $>$ 60	0.262	0.0370	0.255	−0.232
Female	−0.105	0.00259	−0.151	0.294*
Education: Above master's degree	−0.300**	−0.0208	−0.317	0.450*
Years of employment	−0.0114*	0.0136*	0.00450	−0.00573
Type of contract ref.: Permanent				
Annual	0.457*	0.134	1.021***	−0.888***
Less than one year	−0.433	0.264	0.949***	−1.472***
Teaching practices				
ICT use	0.0862	−0.0989	−0.111	0.0909
Flipped classroom ref.: Never				
Sometimes	−0.135	0.110	0.0815	−0.403**
Often or always	0.300*	−0.265	0.101	−0.413
Peer activities ref.: Never				
Sometimes	0.0938	−0.200	0.122	0.0956
Often	−0.213	0.346	−0.130	0.273
Always	−0.185	0.323	−0.0515	0.0830
Classroom presentation ref.: Never				
Sometimes	−0.0810	−0.0691	−0.0421	−0.198
Often or always	−0.0159	−0.301	−0.254	0.344
Laboratory activities ref.: Never				

Table 8 (continued)

	Reading		Mathematics	
	Under- achievement	Over- achievement	Under-achievement	Over-achievement
Sometimes	−0.0845	0.180	0.0368	−0.0929
Often or always	0.00368	0.171	−0.000435	−0.0841
Group work evaluation ref.: Never				
Sometimes	0.131	−0.161	0.313**	−0.203
Often or always	0.184	−0.0637	0.419*	−0.385
ICT for testing and evaluation ref.: Never				
Sometimes	0.0892	−0.0868	0.277*	0.0355
Often or always	0.0384	−0.108	0.159	0.212
Type of class ref.: Only in-person				
Primarily in-person	0.0550	−0.187	0.0180	−0.0434
Primarily remote	0.716	−0.773***	−0.598	0.110
Equally in-person and remote	0.203	−0.358	0.420	−1.199***

Table 9 Intraclass Correlation Coefficients (ICC, %) across model specifications

	Reading		Mathematics	
	Under-achievement	Overachievement	Underachievement	Over-achievement
Null model	44.0	34.3	50.2	53.4
Student covariates	15.9	11.7	27.2	26.3
Full model	13.9	9.3	24.7	21.4

ICCs are reported as percentages and are computed assuming the residual individual level variance equal to $\pi^{2/3}$

dichotomised items that make up the scale. No coefficients were found to be significantly different from zero.

Lastly, data from the teacher questionnaire shed light on the impact of remote teaching, still in use due to COVID-19 restrictions at the time of data collection. While different degrees of remote learning have been found to be not significantly associated with under-achievement, they show to reduce the odds of over-achieving in both subjects. Students primarily taught remotely have 54% lower odds of over-achieving in reading compared to peers who attended in-person classes. Similarly, students with a balanced mix of remote and in-person classes were found to be three times less likely to over-achiever in mathematics.

Table 9 reports the Intraclass Correlation Coefficients (ICC) for three different specifications of the multilevel model: i) an empty model with only a random intercept at the classroom level to understand variance decomposition; ii) a model with only student-level covariates; and iii) the full specified model. The variation of the ICC across models with the same outcome variables shows how the unexplained portion of the variance at the class level changes when additional covariates are introduced into the model. The first row of the table shows that a significant portion of the variance in the probabilities of under-achieving and over-achieving in both subjects is attributable to unobserved characteristics within the classrooms. This is more evident for mathematics achievement (consistently over 50%) than for reading (44% for under-achievement and 34.3% for over-achievement). Introducing student-level covariates produces a steep reduction in unexplained classroom-level variance (around 30 percentage points for reading and around 25 for mathematics). The addition of teacher-level covariates produce a further, though limited, reduction of the unexplained classroom-level variance. The remaining ICCs in the full model are still non-negligible, especially in mathematics, pointing to the persistence of relevant unobserved classroom-level characteristics associated with the probabilities of our outcomes.

5 Conclusions

The present contribution introduces several novel aspects to the existing literature on factors which are related with divergences in academic success and failure monitored in terms of over-or-under achieving. Firstly, it simultaneously considers competences in two skill areas, reading and mathematics, that have traditionally been widely investigated separately in large-scale assessment research. By integrating results on both these dimensions, the study provides robust and comprehensive insights that can inform the design of future policies aimed at enhancing school effectiveness and promoting equitable educational outcomes across diverse student populations.

The main findings provide evidence in response to the research questions, highlighting how teachers' characteristics, instructional practices, and ICT-related factors are associated with students' attainment. With respect to the first RQ1 (concerning the relation between teachers characteristics and students achievement), the main findings confirm that teachers' experience, qualifications, and employment conditions are significantly associated with student performance, although with differences across subjects. Teaching experience is positively related to reading outcomes, as it reduces the probability of under-achievement and increases the likelihood of over-achievement. Teachers with a high qualification above a master's degree is also associated with a lower probability of under-achievement in reading and a higher likelihood of over-achievement in mathematics. In addition, teacher gender is significantly associated with mathematics, where female teachers are linked to a higher probability of student over-achievement. Employment conditions emerge as a relevant factor: temporary contracts are associated with higher risks of under-achievement and lower probabilities of over-achievement in mathematics, and with increased under-achievement in reading. These results underline the importance of the contract stability for teachers and advanced training, while also showing that their effects vary across tracks.

Turning to the second RQ2 (Teaching practices and student success/failure), the results indicate that instructional practices are only weakly associated with over-achievement, as none of the considered practices show significant positive effects on over-achievement. However, some practices are associated with a higher likelihood of under-achievement. Specifically, more frequent use of flipped classrooms is linked to an increased probability of under-achievement in reading, while in mathematics it is associated with a lower likelihood of over-achievement. Similarly, group work assessment practices are positively associated with under-achievement in mathematics, especially when used more frequently. These findings suggest that such practices may not effectively support lower-achieving students.

Finally, regarding the role of ICT (RQ2.1), the results do not provide evidence of a significant overall effect of ICT use on student achievement. The general ICT use scale is not associated with either under- or over-achievement in reading and mathematics. However, specific ICT-related practices show some negative associations: ICT-based assessment is linked to a higher probability of under-achievement in mathematics. In addition, remote teaching is associated with a lower likelihood of over-achievement in reading, while a mixed approach of instruction (both in-person and remote) reduces the probability of over-achievement in mathematics. Overall, these results suggest that ICT does not act as a facilitator of achievement and may be associated with lower performance. Summarising, teachers' characteristics such as experience, qualifications, and employment stability, play a relevant role in shaping student achievement compared to the instructional practices here investigated and ICT use, whose effects appear more limited and, in some cases, negatively associated with performance.

This study provides a starting point for further investigation into the differential effects of educational practices on student competencies across various high school tracks. Future research could build on these findings by employing causal inference methods, such as propensity score matching, to more precisely isolate the impact of specific instructional practices on student outcomes and to deepen our understanding of the mechanisms shaping academic achievement. This approach was not explored here as its main aim was to detect how teaching practices at class level, teacher's assets in terms of experience and qualification and students' profiles are associated with academic success or failure, both monitored in terms of level of competences achieved, carrying out the analysis at student and class level. In this framework, the presence of first level units clustered in classes and the importance of key covariates at class level, rises caution to the use of propensity score methods to match students with similar profiles. Ignoring the presence the Level 2 units in the matching procedure could lead to serious biased estimates as cluster level characteristics are partially lost when matching is carried out only at level 1 units.

Nonetheless, it is worthwhile to rely on multiple waves of the INVALSI surveys in order to assess the robustness of these results across years. Furthermore, the teachers' survey do not cover all classes investigated by the students' survey and this structural missingness in the master data set subsequent merging with students' data can not be easily recovered with imputation procedures, without risking of introducing other sources of bias. Lastly, the data do not provide key information on factors such as students' level of digital skills or teachers' training in the use of ICT for teaching, neither on their training in specific instructional practices used in the classroom. All these aspects may influence the observed results and should be considered in future research lines (Table 14).

Table 10 Missing data in the whole sample ($n = 11{,}169$)

Variable	Missing (%)
Migratory background	1.2
Reading score	6.7
Mathematics score	6.9
Literature teacher data	25.5
Mathematics teacher data	26.8

Table 11 Distribution of missing cases in the whole sample ($n = 11{,}169$) by school track and geographical area

Category	Missing cases (%)
School track	
Other general tracks	14.3
Vocational institute	14.9
Technical institute	27.9
Classical and linguistic general track	20.9
Scientific general track	22.0
Geographical area	
Centre	22.5
North	37.0
South and Islands	40.5

Table 12 Percentage of students exposed to teaching practices during reading class (n=8,320) and mathematics class (n=8,171), by track

	Scientific Gen. Track		Classic/ Linguistic Gen. Track		Other Gen. Track		Technical		Vocational		Total	
	Read	Math	Read	Math	Read	Math	Read	Math	Read	Math	Read	Math
Peer activities (%)												
Never or almost never	16.7	29.9	17.3	23.1	15.7	19.9	10.9	20.1	5.6	17.0	13.1	22.4
Sometimes	52.4	37.2	35.7	39.5	50.3	44.2	41.0	46.8	36.8	36.1	43.8	41.5
Often	22.9	25.3	30.7	35.6	20.8	35.0	39.4	27.0	47.3	36.1	32.5	30.3
Always or almost always	8.1	7.6	16.3	1.9	13.2	0.9	8.7	6.1	10.3	10.7	10.6	5.8
Classroom presentations (%)												
Never or almost never	15.6	36.1	13.3	23.6	15.4	28.6	4.9	38.0	6.6	31.9	10.5	33.3
Sometimes	50.3	42.2	42.9	50.3	39.8	50.1	47.5	47.8	31.4	47.5	44.0	47.1
Often	27.6	21.7	32.1	19.8	31.0	21.3	42.8	11.3	47.7	20.6	36.5	17.8
Always or almost always	6.6	0.0	11.7	6.3	13.9	0.0	4.8	2.9	14.3	0.0	9.0	1.9
Flipped classroom (%)												
Never or almost never	43.5	51.0	33.8	41.9	26.9	47.2	27.0	39.8	19.0	35.7	30.8	43.3
Sometimes	38.3	32.3	47.5	49.0	50.6	49.5	49.5	51.1	53.6	46.5	47.2	45.3
Often	16.9	16.6	15.7	9.0	21.0	3.3	21.9	8.2	25.9	16.0	20.2	10.9
Always or almost always	1.3	0.0	3.1	0.0	1.6	0.0	1.6	0.9	1.5	1.9	1.7	0.6
Laboratorial activities (%)												
Never or almost never	33.1	22.2	33.8	34.4	30.5	35.4	22.9	46.5	18.8	24.1	27.4	34.0
Sometimes	45.0	58.9	36.8	45.3	46.8	54.5	50.7	44.5	63.7	43.1	48.7	49.3
Often	18.0	17.6	27.0	17.4	19.4	9.2	24.2	8.9	14.3	28.5	20.9	15.2
Always or almost always	4.0	1.3	2.4	2.9	3.4	0.8	2.2	0.0	3.2	4.3	3.0	1.5
Type of class (%)												
Mostly in person	72.6	73.1	77.7	82.7	70.0	72.7	76.0	72.0	80.2	65.9	75.1	73.0
Mostly remote	0.0	0.4	0.0	0.0	1.6	0.0	0.7	1.0	0.6	0.0	0.5	0.4
Equally in person and remote	4.4	1.6	0.0	3.3	2.4	0.8	1.3	1.9	1.7	6.1	2.1	2.5
Only in person	23.0	24.9	22.3	14.0	26.0	26.5	22.1	25.2	17.5	28.1	22.2	24.1

Table 13 Percentage of students exposed to evaluation strategies in reading (n=8320) and mathematics (n=8131), by track

	Scientific Gen. Track		Classic/ Linguistic Gen. Track		Other Gen. Track		Technical		Vocational		Total	
	Read	Math	Read	Math	Read	Math	Read	Math	Read	Math	Read	Math
Group evaluations (%)												
Never or almost never	32.9	53.6	33.0	33.4	31.5	32.5	18.8	52.7	13.7	37.0	25.4	45.1
Sometimes	55.9	38.7	47.1	53.8	50.3	63.6	59.1	38.9	75.3	48.5	57.7	45.8
Often	9.9	7.7	18.3	12.8	15.6	3.9	18.4	8.2	11.0	14.5	14.8	9.1
Always or almost always	1.3	0.0	1.7	0.0	2.6	0.0	3.7	0.2	0.0	0.0	2.1	0.1
ICT for evaluation (%)												
Never or almost never	55.7	60.5	77.7	82.7	70.0	72.7	76.0	72.0	80.2	65.9	75.1	73.0
Sometimes	37.1	34.5	0.0	0.0	1.6	0.0	0.7	1.0	0.6	0.0	0.5	0.4
Often	5.9	5.1	0.0	3.3	2.4	0.8	1.3	1.9	1.7	6.1	2.1	2.5
Always or almost always	1.3	0.0	22.3	14.0	26.0	26.5	22.1	25.2	17.5	28.1	22.2	24.1

Table 14 Estimated coefficients (log odds) for the students covariates only model on students' probability of under-achievement and over-achievement in Reading and Mathematics

	Reading		Mathematics	
	Under-achievement	Over-achievement	Under-achievement	Over-achievement
Intercept	−0.943***	−2.684***	−0.610***	−2.006***
Student characteristics				
ESCS	−0.169***	0.309***	−0.0494	0.178***
School track *ref.: Other general track*				
Vocational institute	1.673***	−2.801***	1.202***	−2.532***
Technical institute	0.366***	−0.619***	−0.605***	0.436*
Classical/Linguistic general track	−1.420***	1.158***	−0.983***	0.650**
Scientific general track	−1.660***	1.262***	−2.913***	2.747***
Female	−0.205***	0.186**	0.568***	−0.774***
Migratory background ref.: Native				
1st generation	0.362**	−1.727***	0.526***	−0.334
2nd generation	0.520***	−0.478*	0.283**	−0.265
Region ref.: North				
Centre	0.802***	−0.672***	0.981***	−0.822***
South	1.506***	−1.116***	1.926***	−1.984***

Tabular appendix

Acknowledgements The paper uses elementary data from INVALSI surveys 2021-2022. The opinion expressed in the works and the responsibility for the conclusions drawn lies solely with the author(s) and does not necessarily reflect the views of INVALSI Institute.

Author contributions NB and IM wrote together the code to perform the statistical analysis described. NB produced table and plots (Figure 1) for the IRT models. Both wrote the first draft of the manuscript. IS and MPV developed the analysis plan, commented and revised the entire manuscript. All authors reviewed the manuscript.

Funding Open access funding provided by Università degli Studi di Milano - Bicocca within the CRUI-CARE Agreement. The paper is financially supported under the National Recovery and Resilience Plan (NRRP), Mission 4, Component 2, Investment 1.1, Call for tender No. 104 published on 2.2.2022 by the Italian Ministry of University and Research (MUR), funded by the European Union – NextGenerationEU – Project Title From high school to university: Assessing peers; influence in educational inequalities and performances – CUP F53D23006150006 - Grant Assignment Decree No. 1060 adopted on 07/17/2023 by the Italian Ministry of Ministry of University and Research (MUR).

Data availability No datasets were generated or analysed during the current study.

Declarations

Conflict of interest The authors declare that they have no conflict of interest.

References

Abukari, M.A., Najah, A.N., Samari, J.A., Gonyalug, I.A., Agyei, P.: Improving teaching and learning of organic chemistry in senior high schools using collaborative approaches. Universal J. Educ. Res. **2**(4), 334–350 (2023). https://doi.org/10.17613/nsgs-tn43

Arifin, Z., Sukarmin, S., Kamari, A.: The effect of inquiry-based learning on students' critical thinking skills in science education: a systematic review and meta-analysis. Eurasia J. Math. Sci. Technol. Educ. **21**(3), em2592 (2025). https://doi.org/10.29333/ejmste/15988

Baker, F.B., Kim, S.H.: Item response theory: parameter estimation techniques. CRC Press, Boca Raton (2004)

Barone, C.: A new look at schooling inequalities in italy and their trends over time. Res. Soc. Stratif. Mobility **27**(2), 92–109 (2009). https://doi.org/10.1016/j.rssm.2009.04.001

Bilici, S., Yilmaz, R.M.: The effects of using collaborative digital storytelling on academic achievement and skill development in biology education. Educ. Inf. Technol. **29**, 20243–20266 (2024). https://doi.org/10.1007/s10639-024-12638-7

Birnbaum, A.: Statistical theory for logistic mental test models with a prior distribution of ability. J. Math. Psychol. **6**(2), 258–276 (1969). https://doi.org/10.1016/0022-2496(69)90005-4

Burroughs, N., Gardner, J., Lee, Y., Guo, S., Touitou, I., Jansen, K., Schmidt, W.: A review of the literature on teacher effectiveness and student outcomes. In: Teaching for excellence and equity: analyzing teacher characteristics, behaviors and student outcomes with timss, pp. 7–17. Springer International Publishing, Cham (2019)

Chinelo, B.O.: Evaluating the effect of activity based method of teaching mathematics on Nigerian secondary school students achievement in mathematics. Puissant (2020). https://doi.org/10.48550/arXiv.2011.10785

Coleman, J.S. , Campbell, E.Q. , Hobson, C.J. , McPartland, J. , Mood, A.M. , Weinfeld, F.D. , York, R.L.: Equality of educational opportunity (Tech. Rep. No. OE 38001). Washington, DC: U.S. Department of Health, Education, and Welfare, Office of Education (1966)

Contini, D., Cugnata, F., Scagni, A.: Social selection in higher education: enrolment, dropout, and timely degree attainment in Italy. High. Educ. **75**(5), 785–808 (2018). https://doi.org/10.1007/s10734-017-0170-9

Contini, D., Di Tommaso, M.L., Mendolia, S.: The gender gap in mathematics achievement: evidence from Italian data. Econ. Educ. Rev. **58**, 32–42 (2017). https://doi.org/10.1016/j.econedurev.2017.03.001

De Ayala, R.J.: The theory and practice of item response theory. Guilford Publications (2013)

Fabrizi, E. , Sulis, I. , Busetta, A. , Ragozini, G.: Intergenerational transmission of disadvantages in the italian labour market. Socio-Economic Planning Sciences **96** https://doi.org/10.1016/j.seps.2024.102097 (2024)

Goldstein, H.: Multilevel statistical models. Wiley & Sons, Chichester (2011)

Goldstein, H., Spiegelhalter, D.J.: League tables and their limitations: statistical issues in comparisons of institutional performance. J. R. Stat. Soc. Ser. A Stat. Soc. **159**(3), 385–409 (1996)

Hambleton, R.K., Swaminathan, H., Rogers, H.J.: Fundamentals of item response theory, vol. 2. Sage, Newbury Park (1991)

Hew, K.F., Bai, S., Huang, W., Dawson, P., Du, J., Huang, G., Thankrit, K.: On the use of flipped classroom across various disciplines: insights from a second-order meta-analysis. Australas. J. Educ. Technol. **37**(2), 132–151 (2021). https://doi.org/10.14742/ajet.6475

INVALSI (2016) Rilevazioni nazionali degli apprendimenti 2015–16. rapporto tecnico (Tech. Rep.). Roma:INVALSI

INVALSI (2019) Rapporto prove invalsi 2019 (Tech. Rep.). Istituto Nazionale per la Valutazione del Sistema di Istruzione

INVALSI (2024) Rapporto prove invalsi 2024 (Tech. Rep.). Istituto Nazionale per la Valutazione del Sistema di Istruzione

Jerrim, J., Oliver, M., Sims, S.: The relationship between inquiry-based teaching and students' achievement: new evidence from a longitudinal pisa study in England. Learn. Instr. **80**, 101310 (2022). https://doi.org/10.1016/j.learninstruc.2020.101310

Kaçar, T., Terzi, R., Arıkan, İ, Kırıkçı, A.C.: The effect of inquiry-based learning on academic success: a meta-analysis study. Int. J. Educ. Literacy Stud. **9**(2), 15–23 (2021). https://doi.org/10.7575/aiac.ijels.v.9n.2p.15

Kyndt, E., Raes, E., Lismont, B., Timmers, F., Cascallar, E., Dochy, F.: A meta-analysis of the effects of face-to-face cooperative learning. do recent studies falsify or verify earlier findings? Educ. Res. Rev. **10**, 133–149 (2013). https://doi.org/10.1016/j.edurev.2013.02.002

Leckie, G., Goldstein, H.: The importance of adjusting for pupil background in school value-added models: a study of progress 8 and school accountability in England. Br. Edu. Res. J. **45**(3), 518–537 (2019). https://doi.org/10.1002/berj.3511

Lo Cicero, A.: School performance gaps in italian regions: Estimating the impact of individual, school, and territorial factors. invalsi 2021/22 data analysis. Working Papers Dipartimento di Scienze Sociali ed economiche 13, (2024)

Pensiero, N. , Giancola, O. , Barone, C.: Socioeconomic inequality and student outcomes in italy. L. Volante, S. Schnepf, J. Jerrim, D. Klinger (Eds), Socioeconomic inequality and student outcomes (pp. 81–94). Singapore:Springer (2019)

Porcu, M., Sulis, I., Usala, C., Giambona, F.: Will the gap ever be bridged? a cross-national comparison of non-native students' educational achievements. Genus **79**(1), 19 (2023). https://doi.org/10.1186/s41118-023-00199-5

Rasch, G.: Probabilistic models for some intelligence and attainment tests. University of Chicago Press, Chicago (1980)

Strelan, P., Osborn, A., Palmer, E.: The flipped classroom: a meta-analysis of effects on student performance across disciplines and education levels. Educ. Res. Rev. **30**, 100314 (2020). https://doi.org/10.1016/j.edurev.2020.100314

Sulis, I., Giambona, F., Porcu, M.: Adjusted indicators of quality and equity for monitoring the education systems over time. insights on eu15 countries from pisa surveys. Socioecon. Plann. Sci. **69**, 100714 (2020). https://doi.org/10.1016/j.seps.2019.05.005

Tocchioni, V. , Milone, S. , Lombardi, G.: Gender disparities in school-to-university transition in italy: the role played by the socio-economic condition and the type of high school. RIVISTA ITALIANA DI ECONOMIA, DEMOGRAFIA E STATISTICA 67–78, https://doi.org/10.71014/sieds.v79i1.351 (2025)

Tomaszewski, W., Xiang, N., Huang, Y., Western, M., McCourt, B., McCarthy, I.: The impact of effective teaching practices on academic achievement when mediated by student engagement: evidence from Australian high schools. Educ. Sci. **12**(5), 358 (2022). https://doi.org/10.3390/educsci12050358

Villanea, G.: Influence of inquiry-based science activities on students' achievement. Psychol. Educ. Multidiscip. J. **16**(1), 45–57 (2023). https://doi.org/10.5281/zenodo.10438669

Yaşar, M.D., Erdoğan, M., Batdı, V., Cinkara, Ü.: Evaluation of cooperative learning in science education: a mixed-meta method study. Eur. J. Sci. Math. Educ. **12**(3), 411–427 (2024). https://doi.org/10.30935/scimath/14872

Yu, R., Singh, K.: Teacher support, instructional practices, student motivation, and mathematics achievement in high school. J. Educ. Res. **111**(1), 81–94 (2018). https://doi.org/10.1080/00220671.2016.1204260

Zheng, L., Bhagat, K.K., Zhen, Y., Zhang, X.: The effectiveness of the flipped classroom on students' learning achievement and learning motivation: a meta-analysis. Educ. Technol. Soc. **23**(1), 1–15 (2020)