Good or Bad? Understanding the Effects Over Time of Multigrading on Child Achievement

Barbetta, G.P.; Chuard-Keller, P.; Sorrenti, G.; Turati, G.

DOI
10.1016/j.econedurev.2023.102442

Publication date
2023

Document Version
Final published version

Published in
Economics of Education Review

License
CC BY

Link to publication

Citation for published version (APA):
Barbetta, G. P., Chuard-Keller, P., Sorrenti, G., & Turati, G. (2023). Good or Bad? Understanding the Effects Over Time of Multigrading on Child Achievement. *Economics of Education Review*, *96*, Article 102442. https://doi.org/10.1016/j.econedurev.2023.102442

Economics of Education Review 96 (2023) 102442

Contents lists available at ScienceDirect

Economics of Education Review

journal homepage: www.elsevier.com/locate/econedurev

Good or bad? Understanding the effects over time of multigrading on child achievement☆

Gian Paolo Barbetta [a], Patrick Chuard-Keller [b], Giuseppe Sorrenti [c,*], Gilberto Turati [d]

[a] Università Cattolica del Sacro Cuore, Milan, Italy
[b] University of St.Gallen, St.Gallen, Switzerland
[c] University of Amsterdam, Tinbergen Institute, and CESifo, Amsterdam, Netherlands
[d] Università Cattolica del Sacro Cuore, Rome, Italy

ARTICLE INFO

JEL classification:
I26
I28
R53
H52

ABSTRACT

Multigrading represents the practice of mixing children of different ages in the same classroom. This paper examines the effect of attending a multigrade class in Grade 2 on students' academic achievement in Grades 2, 5, and 8, respectively, considering Italy as a case study. To address the issue of endogeneity of multigrading (and class size), we adopt an IV identification strategy based on a law that disciplines class composition. We show that multigrading has a positive (16 percent of a standard deviation) short-term effect on academic achievements. However, this effect diminishes over time and becomes negative (-10 percent of a standard deviation) if students spend several years in a multigrade class. Mechanism analysis indicates the fundamental role of teachers and suggests that the negative long-term effect of multigrading is not statistically different from zero when multigrade classes are taught by more experienced teachers. These findings, based on longitudinal data, reconcile contrasting results in the literature, which are based on cross-sectional data and on the short-term effects of multigrading.

1. Introduction

At the turn of the century, UNESCO (2004) estimated that about one-third of all children worldwide attended multigrade classes, i.e. classes that mix children of different ages in the same classroom. Although more recent estimates are unavailable, e.g., Cornish (2021) suggests that "multigrade schooling is ubiquitous throughout the world but the context of multigrade schooling varies from country to country". In light of this variability, understanding the effect of multigrading on child development becomes important for at least two reasons: first, to ascertain the impact of this teaching practice in isolated areas where multigrade classes are a necessity, due to constrained economic and human resources; second, to understand the consequences of this practice in areas where it has gained support even if the circumstances would allow for single-grade classes.

Despite its widespread use, the evidence of the *causal* impact of multigrading on child development is mixed, mostly based on cross-sectional data, and almost exclusively focused on short-term outcomes.[1] This paper tries to overcome these limits by assessing both the short-term and the cumulative[2] impacts of multigrading on children's cognitive achievements using a new longitudinal data set that contains repeated test scores for Italian students in primary and lower-secondary education. Specifically, for the first time since the introduction of the standardized INVALSI (National Institute for the Evaluation of the Instruction and Training System) tests in Italy,[3] we are able to follow a cohort of students over time and record their math and language

☆ We wish to thank Lorenzo Rocco, two anonymous reviewers, and participants at IWAEE 2021 for helpful comments. We also thank Lucia De Fabrizio (MIUR) for kindly providing data about Italian schools and teachers, and Francesco Porcelli for making available SOSE data.

* Corresponding author.
E-mail address: g.sorrenti@uva.nl (G. Sorrenti).

[1] Earlier studies surveyed in Little (2001) investigate the effect of multigrading. However, they do not properly address the endogenous sorting of students into multigrade classes. Leuven and Rønning (2016), Checchi and De Paola (2018), and Barbetta et al. (2021) analyze the causal impact of multigrading for students in primary and secondary schools using cross-sectional data.

[2] We label as cumulative those outcomes measured more than five years after attendance of a multigrade class.

[3] The INVALSI test was introduced by Law 176/2007 and is now administered annually to monitoring the skills in mathematics and language of students attending Grades 2, 5, 8, 10, and 13. Each test includes a set of multiple-choice items followed by open-response questions. Students must complete the test in 45 to 90 minutes, depending on the subject and grade.

https://doi.org/10.1016/j.econedurev.2023.102442
Received 6 August 2022; Received in revised form 13 July 2023; Accepted 15 July 2023
Available online 12 August 2023

test scores in Grade 2 and 5 of primary school, and in Grade 8 of lower-secondary (or middle) school.[4] Thanks to these data we can compare the development of the cognitive skills of students attending, respectively, single- and multigrade classes in Grade 2.

We follow Barbetta et al. (2021) to deal with endogenous sorting into multigrade classes (and class size): we use an instrumental variable (IV) identification strategy, which exploits a regulation that prescribes different cut-offs, defined out of the number of students in a specific cohort, to discipline both the formation of single- versus multigrade classes and class size.

Our results show that attending a multigrade class in Grade 2 positively affects the short-term performance – about 16 percent of a standard deviation, statistically significant at the 1 percent level – in the standardized test in math and language. However, in Grade 5 the effect declines to zero, while in Grade 8 it turns into a negative and statistically significant 10 percent of a standard deviation. The findings are robust across different empirical specifications.

We then consider whether the long-term cumulative detrimental effect of multigrading could be avoided. We look at both teachers and resources. The analysis of the mechanisms suggests that "how" multigrade teaching is implemented by teachers is the key driver for the success of this practice. Specifically, when students attend multigrade classes in Grade 2 in schools with a large share of teaching personnel under permanent contracts, the long-term detrimental impact of multigrading is reduced by half (−5 percent) and is not statistically different from zero. Conversely, when students attend multigrade classes with a high share of temporary teachers, the negative long-term effect of multigrading persists (−16 percent) and is statistically significant at the usual levels. This evidence suggests that there exist ways of implementing multigrade teaching without negative consequences on children's cognitive development, including in contexts with strong persistence of multigrading.[5]

In addition, we show that attending a multigrade class in primary school does not increase students' probability of enrolling in worse lower-secondary schools, defined in terms of school resources. Therefore, this channel is unlikely to explain the cumulative effect of multigrading.

Our results contribute to the literature in two main ways. First, our estimates based on longitudinal data help understanding some puzzling results in the existing literature, exclusively based on cross-sectional data: the positive impacts for younger cohorts found by Leuven and Rønning (2016) and Barbetta et al. (2021) versus the negative impacts for older cohorts in Grade 5 found by Checchi and De Paola (2018) and Sims (2008). Given the strong persistence of multigrading, our findings, based on longitudinal data, show that multigrading positively affects achievement in the early stages of the school career. However, the cumulative effect turns negative after some years. Second, we show that the possible negative effects of multigrading could be avoided by investing in effective teaching practices. The policy implication of the latter result is crucial for students living in scarcely populated areas where multigrading is the only available (and possible) schooling option.

[4] The analysis of Grade 8 is particularly important as, upon its completion, students are tracked into different secondary schools. The school performance in Grade 8 plays a major role in the choice of the school track. Different secondary school tracks are associated with different probabilities to access university and different future labor market returns.

[5] Admittedly, our analysis might mask some correlation between teachers' characteristics (and type of contract) and other school resources. This limitation does not undermine our analysis as this exercise only aims at understanding whether it is possible to avoid the negative long-term impact of multigrading through the provision of a high-quality standard of multigrade teaching. Whether the effect is only driven by teachers or also other school resources holds second-order importance for the current analysis.

The remainder of the paper is structured as follows. In Section 2, we provide essential background information on the Italian schooling system and the rules governing class formation. Section 3 describes the longitudinal data used in the estimations and Section 4 defines the empirical strategy to identify the impacts of multigrading. Section 5 reports the results of the empirical analysis. Section 6 investigates the mechanisms underlying the multigrade effect. Finally, Section 7 concludes the paper.

2. Institutional background

We focus on students in their first cycle of education, which includes primary and lower-secondary school. In Italy, primary education (ISCED 1) starts at the age of six and lasts five years (Grades 1 to 5), while lower-secondary education (ISCED 2) starts at the age of eleven and lasts three years (Grades 6 to 8). Attendance of the entire first cycle of education is compulsory and free of charge. In Grade 8, students take a national exam to gain a lower-secondary education diploma that gives access to upper-secondary education (ISCED 3), the first two years of which are also compulsory.

In 2021, according to data provided by the National Institute of Statistics (ISTAT), more than 15,000 primary schools operated in the country, with about 2,600,000 students enrolled in Grades 1 through 5. The number of lower-secondary schools is much lower, with roughly 8,000 institutions hosting more than 1.5 million students over the three grades. Unfortunately, no official statistics about multigrading are currently available. However, in public schools (the large majority of schools) this practice is adopted only when a cohort of students is not large enough to create a single-grade class, which happens mostly in inner areas and small municipalities. According to our data, about 23 percent of primary schools located in municipalities with no more than one school adopt multigrade teaching.

Class formation in Italian schools is regulated by law DPR 81/2009. This law defines thresholds – based on the number of students of the same grade enrolled in a specific primary or lower-secondary school – that influence both the probability of being assigned to a multigrade class and class size.

For *primary* schools, law DPR 81/2009 (Article 10) specifically establishes that:

- single-grade classes should enroll a minimum of fifteen and a maximum of twenty-six students;
- multigrade classes should enroll a minimum of eight and a maximum of eighteen students;
- in isolated villages, small islands, and areas characterized by linguistic minorities, single-grade classes could be created with a minimum of ten students. Moreover, the law allows to reduce the maximum number of students when children with disabilities are enrolled in a class.

Rules for primary and lower-secondary schools are very similar. We exploit these rules in Section 4 to build instrumental variables for attending a multigrade and for class-size.

3. Data

We exploit different waves of INVALSI data to build a longitudinal data set on students (our observation unit) who were second-graders in the school year (SY) 2012/2013, following their career in SY 2015/2016, and SY 2018/2019, when they were in Grade 5 and 8, respectively. Data are fully anonymous and only report the numeric "INVALSI code", which identifies each student for her entire school career, plus numeric class and school codes. Neither students, classes, nor schools can be identified.

We consider the INVALSI test scores in math and language of these students in Grades 2, 5, and 8 as the outcome variables in our analysis. Math and language scores are the only available for all

Table 1
Comparison between the Full, the Multigrade, and the Single-grade Samples.

	Full Sample		Multigrade Class		Single-Grade Class		Difference	
	Mean	St. Dev.	Mean	St.Dev.	Mean	St.Dev.	(MG-SG)	t
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Combined Test Score Grade 2	0.00	1.00	0.25	1.03	0.08	0.95	0.17***	(10.18)
Combined Test Score Grade 5	0.00	1.00	0.12	1.01	0.08	0.96	0.04**	(2.54)
Combined Test Score Grade 8	0.00	1.00	−0.04	0.98	0.09	0.98	−0.13***	(−8.15)
Years in Multigrade (MG) Class	n.a.	n.a	4.45	1.45	0.04	0.32	4.41 ***	(502.20)
Class Size Grade 2	20.61	4.27	13.88	2.65	19.73	3.94	−5.85***	(−131.58)
Class Size Grade 5	20.51	4.29	13.07	3.17	19.42	3.83	−6.35***	(−122.16)
Female	0.49	0.50	0.48	0.50	0.50	0.50	−0.02**	(−2.09)
Italian	0.89	0.31	0.93	0.25	0.91	0.28	0.02***	(4.33)
Immigrant 1st Generation	0.03	0.17	0.03	0.16	0.02	0.14	0.01**	(2.36)
Immigrant 2nd Generation	0.07	0.26	0.04	0.20	0.07	0.25	−0.02***	(−7.25)
Father University	0.12	0.32	0.05	0.22	0.08	0.26	−0.03***	(−7.54)
Mother University	0.15	0.35	0.08	0.28	0.11	0.31	−0.02***	(−5.50)
North-West	0.26	0.44	0.44	0.50	0.47	0.50	−0.03***	(−4.20)
North-East	0.19	0.39	0.08	0.28	0.18	0.38	−0.09***	(−20.54)
Center	0.19	0.39	0.12	0.32	0.11	0.31	0.01	(1.52)
South	0.21	0.40	0.29	0.45	0.18	0.38	0.11***	(15.15)
Islands	0.15	0.36	0.07	0.26	0.06	0.24	0.01**	(2.32)
N	505,176		4,033		66,655		70,688	

Notes: This table provides summary statistics for the full sample of students attending Grade 2 in SY 2012/2013 (columns 1 and 2), the sample of students in multigrade (MG) classes (columns 3 and 4), and the sample of students in single-grade (SG) classes (columns 5 and 6). Columns 7 and 8 provide the comparison (difference and t-statistic for the difference) between the sample of students in multigrade classes versus students in single-grade classes. Significance levels: * $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$.

grades. To ease the interpretation of the results, the test scores in math and language are combined to obtain a single performance measure with a mean of zero and standard deviation of one. As covariates, we include background information about students' characteristics – such as gender, nationality and pre-primary school attendance, parental education level, and profession – all provided by INVALSI.

Unfortunately, before SY 2018/2019, INVALSI did not report the grade composition of each class, making it impossible to classify students as attending a single- versus multigrade class. As fully detailed in the Online Appendix (see Section A), we overcome this limitation for students attending institutions located in municipalities hosting only one school.[6] While this slightly limits our analysis, the precise identification of these schools allows us to enrich our data set with additional data on school and municipality characteristics that are important to explore the mechanisms underlying the effect of multigrading (see Section 6).

Our final sample includes the entire cohort of Italian second graders who attended a primary school located in a municipality hosting only one primary school in SY 2012/2013. Overall, we end up with a sample of 70,688 students out of the about 500,000 in each year-cohort, and 3,999 primary schools out of 15,248 registered at the national level.[7]

Table 1 shows summary statistics for the whole population of Italian students attending Grade 2 in SY 2012/13, and for the sub-samples of students who have attended a single-grade (about 94 percent of our sample) and a multigrade class (about six percent) in Grade 2. In the last two columns, we also test the differences in means across the two sub-samples. With respect to the whole population, students in schools included in our sample perform slightly better. Class size is smaller for both single- and multi-grade, but students' background characteristics are similar, despite our sample over-represents schools in the North-Western area of the country where small villages hosting only one school are more present.

When looking at the two sub-samples, the average combined test score shows a decline from Grade 2 to Grade 8 for students in a multigrade class, while it remains stable for those in a single-grade class. Multigrading is persistent over time; students who start in a multigrade class tend to remain in a multigrade class for their entire primary school education. The average class size is clearly larger for single-grade (about nineteen students per class) than for multigrade (about thirteen students). Despite the statistically significant differences in means shown in the last two columns, students and their family background characteristics are fairly similar. Multigrade classes are more present in the South and less in the North-East. We will address the difference in observables between the sample of students in multigrade versus single-grade classes by (i) controlling for all of the observable characteristics displayed in the table in our empirical model; and (ii) dealing with possible unobservable factors shaping the attendance of a multigrade class in an IV setting.

4. Empirical strategy

We aim at estimating the causal effect of attending a multigrade class in Grade 2 on a child's standardized test score in Grades 2, 5, and 8. We compare schools (between-school comparison) with treatment (attendance of a multigrade class) defined as a function of school-specific cohort size.

Our baseline empirical model takes the following form:

$$Score_{i,g\in[2,5,8]} = \beta_0 + \delta MG_{i,g=2} + \beta_1 CS_{i,g=2} + \boldsymbol{X}'_{i,g=2}\boldsymbol{\beta_2} + \epsilon_{i,g=2,5,8} \quad , \qquad (1)$$

where i is an index for each student and g stands for the grade. $Score$ represents the mean of student i's standardized performance in the math and language sections of the INVALSI test taken in grade g. MG is an indicator for student i's attendance of a multigrade class in Grade 2. CS represents the class size in Grade 2. $\boldsymbol{X}$ is a vector containing a set of control variables for child characteristics (age, gender, nationality, first- and second-generation immigrant) and parental background (mother's and father's education and profession). It also includes controls for macro-regions (North-West, North-East, Center, South, Islands). All control variables are measured when the students attend Grade 2 of primary school. ϵ is the error term.

Identifying the causal effect of multigrading on school performance is challenging, as attendance of a multigrade class might correlate with unobserved determinants of a student's achievement. Thus, OLS estimates of the effect of multigrading on children's school performance might suffer from bias due, e.g., to selection on unobservables. This threat is reinforced by the descriptive analysis, which highlighted that

[6] About 18 percent of Italian municipalities do not host any school while 82 percent host at least one primary school. No more than 35 percent of those hosting a primary school host more than one institution; these are quite common only in urban areas.

[7] In Appendix A.3 we discuss external validity concerns implied by this sample selection.

multigrade classes usually include fewer students than single-grade classes. Therefore, a credible empirical analysis needs to isolate the effect of multigrading from the effect of class size.

Following Barbetta et al. (2021), our identification strategy exploits the Italian law DPR 81/2009. The law disciplines how classes should be formed in primary schools, providing an exogenous source of variation to be used to overcome endogeneity concerns relative to MG, the indicator for attendance of a multigrade class in Grade 2. As discussed in Section 2, the law DPR 81/2009 prescribes cut-offs based on the number of students of the same cohort enrolled in the same school. Given that in Italy we never observe schools with both a multigrade and a single-grade class for the same grade, as in our previous work, we construct four mutually-exclusive indicator variables for cohort size to be used as instruments for two variables: (i) the (actual) attendance of a multigrade class and (ii) class size. More precisely, the first indicator ($\mathbb{1}[CohortSize_s < 10]$) takes the value of one if the cohort enrolled in a certain school s contains fewer than ten students. This variable should be a strong predictor of the probability of attending a multigrade class for students in school s. The second indicator ($\mathbb{1}[10 \leq CohortSize_s < 15]$) is for cohorts made by ten to fourteen students and should show a positive correlation with attendance of a multigrade class. The third indicator takes value of one for a cohort with 15 to 26 students ($\mathbb{1}[15 \leq CohortSize_s < 27]$). It should not affect the attendance of a multigrade class, although it should shape class size. Finally, the last indicator ($\mathbb{1}[CohortSize_s \geq 27]$) takes value of one if the cohort size in school s exceeds 26 students and should only explain class size as the cohort should be too large for multigrade classes to be formed.

Considering these four indicator variables for cohort size, the first-stage equation for MG is as follows:

$$\begin{aligned} MG_{i,g=2} &= \lambda_0 + \lambda_1 \mathbb{1}[CohortSize_s < 10] + \lambda_2 \mathbb{1}[10 \leq CohortSize_s < 15] \\ &+ \lambda_3 \mathbb{1}[15 \leq CohortSize_s < 27] + \lambda_4 \mathbb{1}[CohortSize_s \geq 27] \\ &+ \lambda_5 CS_{i,g=2} + \boldsymbol{X}'_{i,g=2}\boldsymbol{\lambda_5} + \delta_{i,g=2} \quad , \end{aligned} \tag{2}$$

where s is the school attended by student i.

We start our analysis considering only MG as endogenous and class size as a standard control variable. But given that class size might suffer from the same sources of endogeneity as attendance of a multigrade class, we also treat class size as an additional endogenous variable in our model. We instrument class size with the set of instruments defined above under the testable assumptions that two of the cut-offs identified by law DPR 81/2009 affect class size while they do not play any role in shaping the probability of observing a multigrade class. The specification that considers class size as endogenous – our preferred one – yields an additional first stage of the following form:

$$\begin{aligned} CS_{i,g=2} &= \tau_0 + \tau_1 \mathbb{1}[CohortSize_s < 10] + \tau_2 \mathbb{1}[10 \leq CohortSize_s < 15] \\ &+ \tau_3 \mathbb{1}[15 \leq CohortSize_s < 27] + \tau_4 \mathbb{1}[CohortSize_s \geq 27] \\ &+ \boldsymbol{X}'_{i,g=2}\boldsymbol{\tau_5} + \mu_{i,g=2} \end{aligned} \tag{3}$$

Section B in the Online Appendix examines the validity of our empirical strategy. First, Table B-1 and Figure B-1 show no differences in students' characteristics around the critical cut-offs identified by the law. Moreover, we discuss identifying assumptions and show that the first-stage estimates (Table B-2) confirm the statistical impact of indicator variables for cohort size on both multi-grade attendance and class size.

5. Multigrading and child achievement over time

This section provides the estimates of the effect of multigrading on child achievement. Table 2 displays the OLS (columns 1 to 3) and the IV second-stage estimates (columns 4 to 9) of Eq. (1). In columns (4) to (6), the model includes class size as a control variable, while in columns (7) to (9) we also consider class size as endogenous. Each specification is replicated for students' test scores in Grades 2, 5, and 8.[8]

The OLS estimates in column (1) suggest that attendance in a multigrade class in Grade 2 is associated with a significant improvement in the test score at the end of the same grade. Students in multigrade classes experience an increase by eight percent of a standard deviation in their Grade 2 test score with respect to second graders attending a single-grade class. This relative advantage disappears in Grade 5 when, as shown in column (2), the test score seems unaffected by attendance of a multigrade class three years earlier. Conversely, multigrading in Grade 2 is associated with a lower performance (−7 percent of a standard deviation) when the test score in Grade 8 is considered as the outcome of interest (column 3).

Columns (4) to (6) display IV results with attendance of a multigrade class considered as the endogenous variable of the model. In column (4), we analyze the short-term impact of attending a multigrade class in Grade 2 on the standardized test score in the same grade. Attendance of a multigrade class causes a statistically significant average increase in the test score by about thirteen percent of a standard deviation. The result is similar to that found by Barbetta et al. (2021). As with the OLS results, the positive effect of multigrading fades away in Grade 5, three years after students' attendance in a multigrade class (column 5). The point estimate for the multigrade effect turns negative in Grade 8, namely at least six years after the first attendance of a multigrade class and once students are in the lower-secondary cycle of education. At this stage of education, the effect of multigrade is statistically significant and accounts for a ten percent of a standard deviation decrease in the standardized test (column 6).

Our findings are robust to considering both attendance of a multigrade class and class size as endogenous variables. Estimates in columns (7) to (9) are similar to those with a single (MG) endogenous variable. Again, the effect of multigrading is strongly positive in the short term and tends to diminish over time, becoming negative in Grade 8.

Our estimates based on longitudinal data help understanding some puzzling results in the existing literature, exclusively based on cross-sectional data, i.e. the positive impacts for younger cohorts showed by Leuven and Rønning (2016) and Barbetta et al. (2021) versus the negative impact for older cohorts in Grade 5 showed by Checchi and De Paola (2018). Our findings highlight the importance of a correct interpretation of the effects – pure versus cumulative – of multigrading in different circumstances, likely explained by the persistence of multigrading over time.

Nonetheless, our analysis leaves an important question unanswered: is it possible – independently of the persistence of multigrading – to adopt forms of multigrade teaching that do not harm long-term child development? The following analyses will try to answer this question.

6. Mechanisms

We identify three potential mechanisms that can help explain our findings from the previous section: (i) the relative age of students in multigrade classes; (ii) effect heterogeneity by "quality" of multigrade teaching; and (iii) the selection of students into different lower-secondary schools after attendance of a multigrade class in primary school. This section only provides suggestive evidence on each of these mechanisms. Indeed, some characteristics of the primary school system, e.g., high persistence of multigrading, prevent us from isolating each of the mechanisms.

[8] All the models consider the combined math-language test score to obtain a comprehensive measure for child cognitive development. Results from Table C-1 demonstrate that the findings remain consistent when considering test scores separately for math and language. Interestingly, the effect appears to be slightly larger for math compared to language.

Table 2
IV Estimates: Multigrading and Educational Performance over Time.

	OLS			IV: CS Exogenous			IV: CS Endogenous		
Dependent Variable:	(1) Score 2	(2) Score 5	(3) Score 8	(4) Score 2	(5) Score 5	(6) Score 8	(7) Score 2	(8) Score 5	(9) Score 8
Multigrading Grade 2	0.080**	−0.005	−0.067***	0.133***	0.007	−0.101***	0.159***	−0.019	−0.094**
	(0.036)	(0.032)	(0.021)	(0.045)	(0.041)	(0.026)	(0.057)	(0.053)	(0.034)
Class Size Grade 2	−0.010***	−0.005	0.000	−0.009***	−0.005	−0.000	−0.005	−0.008	0.001
	(0.003)	(0.003)	(0.002)	(0.003)	(0.003)	(0.002)	(0.006)	(0.006)	(0.004)
Individual Controls	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Parent Controls	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Regional FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
N	70,688	70,688	70,688	70,688	70,688	70,688	70,688	70,688	70,688

Notes: This table shows the OLS and second-stage estimates for the effect of multigrading on students' test scores. The dependent variable is the student's test score in Grade 2 (columns 1,4, and 7), Grade 5 (columns 2, 5, and 8), and Grade 8 (columns 3, 6, and 9). Columns (1) to (3) report OLS estimates. Columns (4) to (6) report IV estimates with only multigrading treated as endogenous variable and class size (CS) included as a standard control variable. Columns (7) to (9) report IV estimates with both multigrading and class size treated as endogenous variables. See text for further details. All models include controls for the child's gender, age, nationality, father's and mother's education and profession, as well as the population and altitude of the municipality. Standard errors are clustered at the school level and reported in brackets. Significance levels: * $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$.

6.1. Relative age and the effect of multigrading

Attending a multigrade class in all grades of the primary school cycle, as shown by descriptive statistics in Table 1, might explain the pattern of the effect of multigrading over time displayed in the empirical analysis. If children ending up in a multigrade class at the beginning of primary school are highly likely to complete the whole primary school cycle in such a class, then these children would switch from being part of the younger cohort in the earlier grades – i.e. Grade 1 and 2 – to belonging to the older one in the classroom in Grade 5. Barbetta et al. (2021) show that being the older in a multi-age class is consistently associated with a poorer performance.

Therefore, the overall effect of multigrading in Grade 5 should be considered as the average *cumulative* effect accrued over the whole primary school cycle of education. Due to dynamic complementarities described in, e.g., Cunha et al. (2010) and Heckman and Mosso (2014), this average cumulative effect carries over in Grade 8, where students experiencing multigrade in all grades at primary school are likely to compare with students of the same age who did not attend a multigrade class in primary school.

Our results do not appear to conflict with those in the relative age literature showing that younger children in same class groups obtain lower mean scores in cognitive tests (Urruticoechea et al., 2021). In fact, in multigrade teaching, children of different ages are in the same classroom but exposed to different school curricula, depending on their school grades. On the contrary, the relative age literature studies the case of children of different ages attending the same school grade and, therefore, exposed to the same curricula. This could reconcile the different results from the two strands of literature suggesting multigrading as a possible educational practice to minimize the negative effects on younger children found by the relative age literature.

6.2. Teachers and teaching practices

Given the negative average long-term impact of multigrading on students' performance, it becomes crucial to understand whether it would be feasible to adopt such a practice without harming the long-term cognitive development of students. To find an answer, we move to school inputs, considering human resources (teachers) in this section, and financial resources, equipment, and school facilities in the next section.

Teachers are among the most important inputs in a child's development process (Alan et al., 2021b; Chetty et al., 2014; Jackson et al., 2014; Xu & Ran, 2020) and their role becomes even more important in contexts of non-standard educational practices such as multigrading. In order to handle classes comprising students of different ages, teachers need to adopt flexible forms of teaching and devote considerable effort to playing their role as educators (INDIRE, 2019, 2020). The National Institute for Documentation, Innovation and Educational Research of the Italian Ministry of Education provides qualitative evidence that teachers' turnover represents one of the main predictors of successful adoption of multigrade classes.[9] A high turnover has multiple possible disadvantages for child development. On the students' side, children – especially those most in need of support by teachers – might become discouraged by continuous changes of their instructors (INDIRE, 2019). On the teachers' side, a high turnover implies that teachers have a shorter time to know the individual characteristics of their students. Moreover, a high turnover may imply that less experienced teachers face the demanding challenge of teaching a multi-age class, and it interrupts the teachers' learning process on how to run a multigrade class effectively. Finally, teachers might feel less motivated and exert a lower effort if they perceive their role as temporary. All of these circumstances might reduce the potential of multigrade teaching in fostering child development.

To investigate the role of teachers' turnover in shaping the effectiveness of multigrading, we match our original data with an additional data set provided by the Ministry of Education and including information about teachers' characteristics at the school level. Table 3 shows the effect of multigrading on child achievement by share of temporary teachers, a proxy of teachers' turnover and teaching experience.[10] To investigate effect heterogeneity, we have performed our IV analysis – with both multigrade and class size treated as endogenous variables – by subsamples of second-grade students. The *low-turnover* subsample includes students attending schools whose share of teachers with temporary contracts is below the median of the distribution. The *high-turnover* subsample includes students attending schools whose share of teachers with temporary contracts is above the median. Columns (1), (3), and (5) report the estimates for the effect of multigrading on children's test scores in Grades 2, 5, and 8, respectively, in schools with low levels of teachers' turnover. Columns (2), (4), and (6) replicate the analysis by focusing on children in schools with high teachers' turnover.

The table conveys an important finding. Teachers' turnover plays a key role in shaping the effectiveness of multigrading. Independently

[9] INDIRE (2019) claims that: *"Teacher turnover has a huge impact: the teacher requires detailed and operational knowledge of the curriculum in order to operate well in a multigrade environment. If the teacher changes, the process starts from scratch every time".*

[10] The average share of temporary teachers in the sample is about 11 percent.

Table 3
Heterogeneity in the Effect of Multigrading by Share of Temporary Teachers.

Dependent Variable:	Score 2		Score 5		Score 8	
	(1)	(2)	(3)	(4)	(5)	(6)
Multigrading Grade 2	0.220**	0.114	0.005	−0.062	−0.047	−0.162***
	(0.089)	(0.080)	(0.081)	(0.078)	(0.049)	(0.049)
Class Size Grade 2	−0.006	−0.004	−0.004	−0.013	0.007	−0.003
	(0.009)	(0.009)	(0.008)	(0.009)	(0.005)	(0.005)
Individual Controls	Yes	Yes	Yes	Yes	Yes	Yes
Parent Controls	Yes	Yes	Yes	Yes	Yes	Yes
Regional FE	Yes	Yes	Yes	Yes	Yes	Yes
Temporary Teacher Share	Low	High	Low	High	Low	High
N	32,972	35,021	32,972	35,021	32,972	35,021

Notes: This table shows whether the multigrade effect is heterogeneous by the share of temporary teachers in the school. The dependent variable is the student's test score in Grade 2 (columns 1 and 2), Grade 5 (columns 3 and 4), and Grade 8 (columns 5 and 6). The sample is split according to the share of temporary teacher in the school during Grade 2 of primary school. *Low* indicates that the primary school has below-the-median temporary teachers. *High* indicates that the primary school has above-the-median temporary teachers. The table report IV estimates with both multigrading and class size treated as endogenous variables. See text for further details. All models include controls for the child's gender, age, nationality, father's and mother's education and profession, as well as the population and altitude of the municipality. Standard errors are clustered at the school level and reported in brackets.
Significance levels: * $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$.

Table 4
Heterogeneity in the Effect of Multigrading by Financial Resources of the Municipality.

Dependent Variable:	Score 8			
	(1)	(2)	(3)	(4)
Multigrading Grade 2	−0.0942*	−0.0933	−0.0984**	−0.0834
	(0.0544)	(0.0581)	(0.0501)	(0.0533)
Class Size Grade 2	−0.00312	0.00842	0.00428	0.00112
	(0.00577)	(0.00625)	(0.00514)	(0.00530)
Individual Controls	Yes	Yes	Yes	Yes
Parent Controls	Yes	Yes	Yes	Yes
Regional FE	Yes	Yes	Yes	Yes
Subgroup	Less Services	More Services	Less Expenditures	More Expenditures
N	30,894	26,091	37,492	25,814

Notes: This table shows the effect of multigrading on students' test scores by municipality resources of the municipality where students attend Grade 8. The dependent variable is the student's test score in Grade 8. The sample is split according to the level of services offered (columns 1 and 2) and expenditures (columns 3 and 4) by the municipality where the student attends Grade 8. *Less* indicates the municipality offer services or report expenditures below the sample median. *High* indicates the municipality offer services or report expenditures above the sample median. The table reports IV estimates with both multigrading and class size treated as endogenous variables. See text for further details. All models include controls for the child's gender, age, nationality, father's and mother's education and profession, as well as the population and altitude of the municipality. Standard errors are clustered at the school level and reported in brackets. Significance levels: * $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$.

of the grade, multigrading always shows a better impact on school performance when teachers' turnover is low. If we consider Grade 8, students who attended a multigrade six years earlier in a school with more permanent teachers perform similarly to their peers attending a single-grade class in a low-turnover school. Interestingly, the negative effect of multigrading in Grade 8 is limited to the subsample of students attending schools with a high share of temporary teachers.

These findings do not change when schools are classified in tertiles of the distribution of the share of temporary teachers (see Table C-2 in the Appendix).

Overall, our analysis suggests that how multigrade teaching is implemented matters. In particular, multigrading could be practiced without harming children's long-term development if teachers' turnover is low. This can imply the selection of more skilled and/or more motivated teachers.

6.3. Multigrading and (future) school resources

The Italian central government defines the allocation and the remuneration of school teachers at the national level and covers their costs.[11] However, for both primary and lower-secondary schools, municipalities have the responsibility to find an appropriate building for hosting the school, bear the costs of maintenance, as well as the cost of some connected services, e.g., a dining hall or school buses. Municipalities differ in terms of available resources, and their choices relative to school buildings and services are also different. In particular, municipalities in rural and remote areas with few inhabitants are generally poorer than less peripheral municipalities. If students attend a multigrade class at the beginning of their educational career and then continue in schools with low educational, financial, and infrastructural resources, this would confound the interpretation of the cumulative effect of multigrading. In this section, we test whether the effect of multigrading in Grade 8 is heterogeneous based on the resources of the school attended after primary education.

As a first step, we merge our data with a new data set provided by SOSE,[12] which allows us to classify each Italian municipality according to the level and quality of services provided to the population and the level of spending.[13] Each municipality is assigned to a *high*- versus a *low-level* group of municipalities according to its relative position with respect to the median of the distribution of services offered in the whole country. The same is replicated for spending. We use this classification to test whether the effect of multigrading changes with the resources of the municipality hosting the school, which are proxied by both the level of services and the level of expenditure.

Column (1) of Table 4 reports the analysis for municipalities with lower (or below-the-median) levels of services. Column (2) analyzes the sample with higher levels of services. Column (3) considers the case of lower-than-median expenditures, while column (4) only includes those municipalities with expenditure above the median of Italian municipalities. As for previous analyses, we estimate IV specifications treating both attendance of a multigrade class and class size as endogenous variables. The table depicts a consistent picture. The restriction to different subsamples based on municipalities' resources does not shape any clear heterogeneity in the multigrade effect. All point estimates are remarkably similar and mimic the point estimates obtained in the whole sample. If anything, the use of subsamples implies an important loss in precision that makes point estimates statistically nonsignificant (or weakly significant).

As a further step, we investigate the heterogeneous effect of multigrading by considering more specific definitions of lower-secondary schools' resources and facilities. To perform this analysis, we use an

[11] The two exceptions are two small Special Statute Regions, Valle d'Aosta and Trentino-Alto Adige, not included in the INVALSI sample. See, e.g., Turati et al. (2017), for additional details.

[12] SOSE is a public company owned by the Italian Ministry of the Economy and Finance and the Bank of Italy that collects an extensive set of information on Italian municipalities. This information is used by SOSE to compute standard needs across municipalities, and to classify municipalities according to their performance with a benchmarking tool called OpenCivitas.

[13] Since we are considering inputs related with financial resources – which vary slowly over time – we use data relative to the SY 2015/2016 in which students in our sample attended Grade 5. This choice allows us to proxy school resources available when students started attending lower-secondary education.

additional source of data by the Ministry of Education that includes information on school resources – such as the availability of rooms with personal computers, a pool/gym or a dining hall – for the whole population of Italian schools. We refer to SY 2018/19, when students were in Grade 8.

We start by considering the share of rooms with a personal computer, which can be considered as a proxy for the digital resources available to schools. We classify schools in three levels, namely *Low*, *Medium*, and *High*, based on the tertiles of the distribution of the share of rooms with a computer in Italian schools. Table C-3 in the Appendix shows that the effect of multigrading is independent of computers' availability.

We then consider the availability of an auditorium, a dining hall, and a sports facility (a gym and/or swimming pool accessible to students) in each school. These three facilities can be considered as proxies of the school environment, and their availability should make the environment livelier and improve students' experience while at school. The estimates reported in Table C-4 in the online Appendix of the paper display limited heterogeneity based on school resources.

Overall, the analysis of municipality spending and school resources does not provide any clear-cut evidence of heterogeneous effects of multigrading. The lack of heterogeneous impacts seems to rule out the possibility that attending a multigrade class shapes the students' probability of ending up in a lower-quality school with fewer financial and non-financial resources. The analysis reinforces the idea that elements more strongly related to teachers' characteristics and the quality of implementation of the educational practice are responsible for the cumulative effects of multigrading.

7. Conclusions

The practice of multigrading is widely used worldwide both to comply with resource constraints and for educational and pedagogical reasons. Despite its use, there remains no consensus on the role of this practice in shaping child development. The few causal studies on multigrading – mostly based on cross-sectional data – draw a mixed picture with some works highlighting a positive effect on students' achievements and others showing a detrimental effect of this practice.

Our study addresses the limitations of previous analyses, which were based on cross-sectional data, by leveraging a new longitudinal data set on cognitive achievements of Italian primary and lower-secondary school students. By doing so, our study contributes to a better understanding of the contrasting findings in the existing literature. We use institutional rules on class formation to deal with the potential endogeneity underlying the attendance of a multigrade class at the beginning of the students' school career. These rules allow us also to isolate the effect of multigrading from the effect of class size.

Our findings show that multigrading in Grade 2 of primary school has beneficial short-term effects on students' cognitive achievements (16 percent of a standard deviation). However, this positive impact fades away over time and becomes negative on average after six years in Grade 8 (about −10 percent of a standard deviation). The persistence of multigrading in the primary cycle of education explains why the effect fades away over time: younger students, who benefit from multigrading in Grade 2, turn older in Grade 5. Hence, the negative effect in Grade 8 is a cumulative effect of an entire primary school career in a multigrade class.

Our results suggest an important policy-relevant question: Is it possible to avoid the negative cumulative effect of multigrading? This question is even more important for those areas where multigrading represents the only possibility for a local school to operate. Investigating the mechanisms, we show that in a context with high persistence of multigrade teaching, the negative cumulative effect of multigrading only appears for those students who attended a multigrade class in a school with a large share of teachers with temporary contracts. These teachers are likely less skilled and less experienced with multigrade teaching. In addition, multigrading is not associated with the choice of future lower-quality schools. Therefore, it seems unlikely that the long-term impact of multigrading is driven by factors such as the quality of schools in lower-secondary education.

Overall, this new evidence suggests that multigrade teaching does not necessarily imply negative cumulative effects on child development. Teachers appear as the most important actors when it comes to guaranteeing the effectiveness of multigrading. As reinforced by anecdotal evidence and teachers' interviews, multigrade teaching requires particular skills, effort, and motivations to be effectively run. Less-motivated and/or less-experienced teachers face too many challenges to deal effectively with students from different grades at the same time.

This study has at least two limitations. First, our results refer to municipalities hosting only one school. Therefore, future research should focus on contexts where the selection process into multigrading might be different, e.g., the case of large cities. Second, our study neglects some potentially important dimensions of the impact of multigrading on a child's development. Due to data limitations, we only focus on students' cognitive development. However, child development is multidimensional, and non-cognitive, behavioral, and socio-emotional skills are also usually deemed as fundamental predictors of future life opportunities (Heckman & Mosso, 2014). Moreover, these soft skills are malleable (Alan et al., 2021a, 2019; Kosse et al., 2020; Sorrenti et al., 2020). Multigrading might play an important role in shaping these skills, e.g., by exposing children to more mature peers or fostering the sense of responsibility toward younger peers by older peers in the classroom. The analysis of the impact of multigrade teaching on a broad and multidimensional set of skills should be prioritized by future research on the topic.

Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Data availability

Data will be made available on request.

Appendix A. Supplementary data

Supplementary material related to this article can be found online at https://doi.org/10.1016/j.econedurev.2023.102442.

References

Alan, S., Baysan, C., Gumren, M., & Kubilay, E. (2021a). Building Social Cohesion in Ethnically Mixed Schools: An Intervention on Perspective Taking. *Quarterly Journal of Economics*, *136*(4), 2147–2194.

Alan, S., Boneva, T., & Ertac, S. (2019). Ever Failed, Try Again, Succeed Better: Results from a Randomized Educational Intervention on Grit. *Quarterly Journal of Economics*, *134*(3), 1121–1162.

Alan, S., Duysak, E., Kubilay, E., & Mumcu, I. (2021b). Social Exclusion and Ethnic Segregation in Schools: The Role of Teacher's Ethnic Prejudice. *The Review of Economics and Statistics*.

Barbetta, G. P., Sorrenti, G., & Turati, G. (2021). Multigrading and Child Achievement. *Journal of Human Resources*, *56*(3), 940–968.

Checchi, D., & De Paola, M. (2018). The Effect of Multigrade Classes on Cognitive and Non-cognitive Skills. Causal Evidence Exploiting Minimum Class Size Rules in Italy. *Economics of Education Review*, *67*, 235–253.

Chetty, R., Friedman, J. N., & Rockoff, J. E. (2014). Measuring the Impacts of Teachers I: Evaluating Bias in Teacher Value-added Estimates. *American Economic Review*, *104*(9), 2593–2632.

Cornish, L. (2021). Introduction. In L. Cornish, & M. J. Taole (Eds.), *Perspectives on Multigrade Teaching* (pp. 1–17). Springer.

Cunha, F., Heckman, J. J., & Schennach, S. M. (2010). Estimating the Technology of Cognitive and Noncognitive Skill Formation. *Econometrica*, *78*(3), 883–931.

Heckman, J. J., & Mosso, S. (2014). The Economics of Human Development and Social Mobility. *Annual Review of Economics*, *6*(1), 689–733.

INDIRE (2019). Tracce di Sostenibilità. Comunità, Pluriclassi e Tecnologie nella Pratica della Scuola. *I Quaderni delle Piccole Scuole, 1/2019*.

INDIRE (2020). Apprendimento Intervallato per Competenze. *I Quaderni delle Piccole Scuole, 7/2020*.

Jackson, C. K., Rockoff, J. E., & Staiger, D. O. (2014). Teacher Effects and Teacher-related Policies. *Annual Review of Economics*, *6*(1), 801–825.

Kosse, F., Deckers, T., Pinger, P., Schildberg-Hörisch, H., & Falk, A. (2020). The Formation of Prosociality: Causal Evidence on the Role of Social Environment. *Journal of Political Economy*, *128*(2), 434–467.

Leuven, E., & Rønning, M. (2016). Classroom Grade Composition and Pupil Achievement. *The Economic Journal*, *126*(593), 1164–1192.

Little, A. W. (2001). Multigrade Teaching: Towards an International Research and Policy Agenda. *International Journal of Educational Development*, *21*(6), 481–497.

Sims, D. (2008). A Strategic Response to Class Size Reduction: Combination Classes and Student Schievement in California. *Journal of Policy Analysis and Management*, *27*, 457–478.

Sorrenti, G., Zölitz, U., Ribeaud, D., & Eisner, M. (2020). The Causal Impact of Socio-emotional Skills Training on Educational Success. *CESifo Working Paper 8197*.

Turati, G., Montolio, D., & Piacenza, M. (2017). Funding and School Accountability: The Importance of Private and Decentralised Public Funding for Pupil Attainment. *Fiscal Studies*, *38*, 615–639.

UNESCO (2004). Educating Rural People: A Low Priority. *The Newsletter of UNESCO Education Sector*, *9*.

Urruticoechea, A., Oliveri, A., Vernazza, E., Giménez-Dasí, M., Martínez-Arias, R., & Martín-Babarro, J. (2021). The Relative Age Effects in Educational Development: A Systematic Review. *International Journal of Environmental Research and Public Health*, *18*(17), 8966.

Xu, D., & Ran, F. X. (2020). The Impacts of Different Types of College Instructors on Students' Academic and Labor Market Outcomes. *Journal of Policy Analysis and Management*, *40*(1), 225–257.