UNIVERSITÀ CATTOLICA DEL SACRO CUORE
Dipartimento di Economia e Finanza

Working Paper Series

Multigrading and Child Achievement

Gian Paolo Barbetta, Giuseppe Sorrenti, Gilberto Turati

Working Paper n. 65

January 2018

UNIVERSITÀ
CATTOLICA
del Sacro Cuore

Multigrading and Child Achievement

Gian Paolo Barbetta
Università Cattolica del Sacro Cuore

Giuseppe Sorrenti
University of Zürich

Gilberto Turati
Università Cattolica del Sacro Cuore

Working Paper n. 65
January 2018

Dipartimento di Economia e Finanza
Università Cattolica del Sacro Cuore
Largo Gemelli 1 - 20123 Milano – Italy
tel: +39.02.7234.2976 - fax: +39.02.7234.2781
e-mail: dip.economiaefinanza@unicatt.it

The Working Paper Series promotes the circulation of research results produced by the members and affiliates of the Dipartimento di Economia e Finanza, with the aim of encouraging their dissemination and discussion. Results may be in a preliminary or advanced stage. The Dipartimento di Economia e Finanza is part of the Dipartimenti e Istituti di Scienze Economiche (DISCE) of the Università Cattolica del Sacro Cuore.

Multigrading and Child Achievement*

Gian Paolo Barbetta† Università Cattolica del Sacro Cuore

Giuseppe Sorrenti‡ University of Zürich

Gilberto Turati§ Università Cattolica del Sacro Cuore

January 2018

Abstract

We exploit Italian law DPR 81/2009, which determines class composition, as an instrument to identify the causal effect of grouping students of different grades into a single class (multigrading) on children cognitive achievement. This article focuses on 7-year-old students—those at the beginning of their formal education. Results suggest that attendance in multigrade classes versus single-grade classes increases students' performance on standardized tests by 15–20 percent of a standard deviation. The positive impact of multigrading only appears for children sharing their class with peers from higher grades and is relatively stronger for students from disadvantaged backgrounds.

JEL classification: I26, I28, R53

Keywords: Multigrade classes, child development, peer effects, rural areas

*First version: November 10, 2017. We wish to thank Gianluca Argentin, Massimiliano Bratti, Claudio Lucifora, and Ulf Zölitz for helpful comments on a preliminary version of this paper, Gianna Barberi (MIUR) for kindly providing data, and Patrizia Falzetti (INVALSI) for both data and comments on a preliminary draft. Chiara Paola Donegani provided excellent research assistance when this project was initiated several years ago. Financial support from the Swiss National Science Foundation (100018_165616) is gratefully acknowledged (Sorrenti). The usual disclaimer applies.

†Università Cattolica del Sacro Cuore, Milan (IT). E-mail: `gianpaolo.barbetta@unicatt.it`

‡University of Zürich (Department of Economics) and Jacobs Center for Productive Youth Development (CH). E-mail: `giuseppe.sorrenti@econ.uzh.ch`

§Università Cattolica del Sacro Cuore, Rome (IT). E-mail: `gilberto.turati@unicatt.it`

1 Introduction

Education in early childhood aims to foster cognitive skills of pupils as well as their individual talents (attentiveness, motivation, self-control, self-confidence) and social traits (positive attitudes toward peers and conflict resolution attitudes) (Cunha and Heckman, 2008, 2010). These skills are largely predictive of future individual success in school, employment, and life in general (Heckman et al., 2006). Therefore, analyzing the effect of programs, practices, or policies that could interact with the development of individual skills during childhood—and early childhood in particular—is a priority (Cunha et al., 2010).

In this work, we study the impact of class composition on the cognitive development of 7-year-old children, focusing on the effect of attending a class with peers of different grades. This practice of mixing more than one grade in a class is referred to as "multigrading". Multigrading is a widespread phenomenon that accounts for about one-third of the total number of classes worldwide according to UNESCO. The practice is particularly common in remote and less affluent areas of the world (such as many areas in developing countries), where its widespread use is often driven by economic constraints, and where children are also more likely to be exposed to fewer learning opportunities. For example, 78 percent of schools in Peru were multigrade in 1998, and multigrading is the only available option for children living in poor, remote areas of Sri Lanka or Vietnam (Hargreaves et al., 2001).

However, multigrading is a common practice in several developed countries as well. Twenty-eight percent of schools in the U.S. reported the use of multiage grouping in 2007; in France, 37 percent of primary school students attend a multigrade class (Leuven and Rønning, 2016). Moreover, multigrade classes account for 70 percent of the classes in Finland and 53 percent in the Netherlands—countries whose students generally achieve excellent scores on international standardized tests such as the OECD PISA test (Mulkeen and Higgins, 2009). In Italy, about 20 percent of schools located in municipalities with no more than one primary school have adopted multigrading.

Multigrade classes have won approval among pedagogists and educational psychologists and have been advanced in several quarters. For instance, they are one of the key ingredients of Montessori schools, which mushroomed in developed countries in an effort to obtain better educational results.[1] Multigrade classes are often proposed as new pedagogical tools that can better adapt to each pupil's rhythm of learning. In Switzerland—starting in 2003 and for the subsequent seven years—the pilot "Basisstufe project" grouped 4- to 8-year-old students in the same class, for about 150 classes and more than three thousand students.[2] In Italy, the project "Piccole scuole" (literally, "Small schools") started by the public National Institute for Innovation and Research in Education (INDIRE) in 2015 helps teachers of multigrade classes in remote areas to work together aided by Information and Communications Technology (ICT). In general, wide anecdotal evidence underlines the potential of multigrading for improving cognitive and noncognitive skills of students.

However, attending a multigrade class could potentially affect child development either positively or negatively via different channels. First, multigrade classes favor interactions among peers who are either more, or less, mature; these interactions could directly foster (for younger peers) or slow (for older peers) the acquisition of cognitive skills, including linguistic and mathematical abilities. At the same time, attending a class with peers of different ages is likely to influence—with unknown results—the noncognitive skills of children, impacting their social and emotional development and influencing friendships and self-perception, as well as other behavioral traits such as altruism or attitudes toward schooling. These personal traits indirectly feed back into achievement. Finally, teaching practices and methods might be influenced by grade composition of the classes, again affecting the children's learning process.

[1] Montessori schools—named after the Italian pedagogist Maria Montessori—consider the use of mixed-age classrooms essential for their educational approach (see, for example, the website of the Association Montessori Internationale). More generally, advocates of multigrade classes say that "the traditional approach of dividing students into single grades based on an arbitrary birth-date range is illogical. [...] Multiage education [...] puts learners at the center, socially and academically. On the social side, younger children look for guidance to older students who know the ropes, while the older students in the classroom organically learn about mentoring, leadership, and collaboration." (The Atlantic, May 9, 2017).

[2] See www.swissinfo.ch/ita/societa/scuola-e-territorio-_pluriclassi–retaggio-del-passato-o-pedagogia-del-futuro-/33582238.

Despite the widespread use and support of multigrading, its effect on students' achievement has been the topic of very few solid empirical studies; evidence of the impact on very young children is even more scarce. Early studies—surveyed, for example, in (Little, 2001)—are unable to properly address sorting of students into multigrade classes. An exception to this literature is the work by Leuven and Rønning (2016), which studies how classroom grade composition affects 15-year-old students' achievements in Norwegian junior high schools. By exploiting a national regulation determining classroom grade composition, the authors show that a one-year exposure to a class that combines two grades increases performance by about 4 percent of a standard deviation. On the contrary, a recent work by Checchi and De Paola (2017) analyzes 10-year-old students in Italy and finds a negative impact of multigrading on standardized tests. However, their work is unable to accurately identify students in multigrade classes. Moreover, they mechanically correct their outcome variable to deal with possible opportunistic and cheating behavior on the test. This correction, as discussed below, is inappropriate for dealing with multigrade classes and, in general, small classes.

Our work complements the existing literature as it represents the first attempt to infer the causal impact of multigrading on early childhood development. We focus here on 7-year-old pupils—children at the very beginning of their schooling career. This is particularly important given that the research by James Heckman and his coauthors has unequivocally shown that early childhood investment in education generates the highest rate of return (see for example Cunha et al., 2010). Most important, by focusing on very young children, we can infer the true impact of multigrading in schools and avoid the cumulative process that is likely to affect older students who might have already experienced multigrading.

We estimate the impact of multigrading on child cognitive development—measured through standardized test scores—by implementing an instrumental variable (IV) approach to address the endogeneity concerns (for example parental preferences) related to attendance in multigrade classes. Moreover, as multigrading is likely to be highly correlated with class size, we separate the effect of class composition from the effect of class size. We focus on Italy and take advantage of DPR 81/2009—a law that regulates the creation of both single-grade and multigrade classes. This law prescribes precise cutoffs, in terms of students enrolled in a specific grade, to establish when a new class should be created. Rules are also established for creating multigrade classes.

We exploit these exogenous cutoffs in a Maimonides' Rule fashion (see Angrist and Lavy, 1999; Angrist et al., 2017) to predict the individual probability of being assigned to a multigrade class. We use this predicted probability as an instrument for the actual grade composition of classes. To tackle the potential correlation between multigrading and class size, we consider two specifications: in the first one, class size is treated as an exogenous control variable; in the second specification, class size (in addition to multigrading) is treated as endogenous and is instrumented with the same cutoffs imposed by DPR 81/2009.

Our outcome variables are represented by the scores of 7-year-old students on the national standardized tests given in all Italian primary schools by the public National Institute for the Evaluation of the Instruction and Training System (INVALSI). The sample of our analysis focuses on students attending school in municipalities where no more than one primary school operates. This restriction depends on data limitations that makes it impossible to identify multigrade classes in municipalities where more than one primary school is active. Nonetheless, this limit mitigates the endogeneity concerns related to school choice and its interaction with parental preferences. In fact, in our sample, parents can only choose one school for their children (the one located in the municipality where they live), unless they are able to afford the costs of commuting to a more distant school in a different municipality.

Consistent with results in Leuven and Rønning (2016) for 15-year-old students, we find that attending a multigrade class at the beginning of primary education positively affects achievements. In our baseline IV model, multigrade attendance increases performance in math and language standardized test score by about 15–20 percent of a standard deviation. This result is robust to: (i) considering class size as an exogenous versus an endogenous variable, (ii) a complete set of sensitivity checks, and (iii) using the time to drive to the closest alternative school as an additional instrument for the attendance of a multigrade class. Moreover, we discuss the possible opportunistic behavior occurring during the administration of standardized test scores in Italy, documented by studies such as Bertoni et al. (2013), Lucifora and Tonello (2015) or Angrist et al. (2017). We show that cheating practices are likely to only marginally affect our point estimates.

We add to the existing literature by finding that the effect of multigrading on child cognitive development appears to be heterogeneous with respect to children's characteristics such as gender and family socio-economic status. We find that females benefit more from multigrading than their male counterparts. More interestingly, we also unveil that children from low socio-economic parental backgrounds (proxied by parental education) obtain higher benefits from multigrading. The latter result suggests that multigrading could mitigate the effect of poor socio-economic conditions on child development, a crucial ingredient for explaining differences in cognitive achievements (see Todd and Wolpin, 2007). Good practices that help children from poor backgrounds are extremely relevant given that child poverty is a massive phenomenon worldwide. In Italy alone, around 1.3 million children (12.5 percent) were living in poverty in 2016 (ISTAT, 2017).

In the last part of our paper we investigate the mechanism underlying the effect of multigrading on cognitive development. We highlight that the positive effect of multigrading on 7-year-old students' achievement is driven by children's sharing the class with more mature peers, namely students from higher grades. On the contrary, multigrading does not show beneficial effects on 7-year-old students attending a multigrade class with younger peers. This result is confirmed when analyzing the performance of students attending the last year of primary schools (fifth graders). These students represent the older cohort in primary schools so they necessarily share a multigrade class with peers from lower grades. We find no effect of multigrading for this last sample of students. This dual evidence suggests that the presence of older peers likely inspires imitation in younger children, therefore improving their performance. Moreover, older peers may increase a younger child's exposure to more refined vocabulary and more advanced topics, which fosters human capital accumulation since an early educational stage.

The remainder of the paper is structured as follows. Section 2 provides essential background information on Italian primary schools and the rules governing class formation. Section 3 describes how we created our data. Section 4 presents our identification strategy. Section 5 discusses our results. Section 6 investigates the mechanism behind our findings. Section 7 concludes.

2 The Institutional Background

Primary school (ISCED 1) in Italy begins for children who are 6-year-old; it covers first to fifth grades. Primary education is compulsory, and its main purpose is to provide sound basic training in reading, writing, and mathematics, plus an elementary understanding of subjects such as geography, history, science, English language, drawing, and music.

Parents can enroll their children in one of the more than 15,000 public primary schools (mostly state-run institutions) or in one of the about 1,500 private schools that operate in the country (according to the 2014 census by the Italian Statistical Office, ISTAT). Public schools enroll more than 93% of the approximately 2.8 million students attending primary school.

No official statistics about multigrading are available. Anecdotal evidence shows that most Italian primary school students attend a single-grade class. According to our data, multigrading is relatively common in Italy, particularly in rural areas and in municipalities where only one school operates.

The estimation of the causal impact of multigrading on individual performance is difficult because students' selection into those classes could be nonrandom. In theory, parents—as well as schools and teachers—could have specific preferences and could therefore try to modify class composition. In practice, this problem should not be very relevant in Italy for two main reasons. First, in creating classes, school principals must follow the rules established by DPR 81/2009. This law defines a set of thresholds to determine when a new single-grade class should be created given the number of students in a single cohort and whether a multigrade class should be formed. Rules are based on the number of students of the same grade enrolled in a specific primary school. The rules specifically establish:

- single-grade classes consist of a minimum of 15 and a maximum of 26 students;
- multigrade classes consist of a minimum of 8 and a maximum of 18 students;
- in special cases—such as isolated villages, small islands, and areas characterized by the presence of linguistic minorities—single-grade classes could be created with a minimum of 10 students. Besides these special cases, the law allows some flexibility (reducing the maximum number of students per class) in the presence of disabled children.

Second, parental preferences are constrained by the specific enrollment process. In fact, public schools adopt uniform criteria to admit students, the main one being the distance between the student's house and the school. Students living in each school catchment area are automatically accepted, but students coming from outside the area can be accepted only if the school has spare capacity. Moreover, national rules require families to apply to a primary school by January-February each year, well before the beginning of the following school year (SY), which starts in mid-September. Within a month of application, school principals are required to communicate to each family whether their children have been accepted. However, students are assigned to classes (and teachers are assigned to each class) only during the summer. Parents cannot participate in this procedure and they only learn of the class composition and the teachers' names shortly before the beginning of the SY (or even the first day of school).

The characteristics of the enrollment process play an important role in our identification strategy (illustrated below) as they mitigate possible selection-into-schools endogeneity concerns. The institutional setting makes it very difficult for parents concerned about grade composition (both in terms of the number of students per class and single versus multigrade class composition) to opt for alternative primary schools. Moreover, public school principals face relevant constraints in exercising their possible preferences on this issue. However, although the effect of parents' and teachers' preferences on class composition appears to be relatively unimportant, we extensively address endogeneity concerns in the remainder of the paper.

3 The Data

Our aim is to compare the educational achievements of children attending either a single or a multigrade class at the beginning of primary school. We measure achievements using individual student scores on the national standardized test run by INVALSI. The INVALSI written test is simply intended to monitor the skills and knowledge of Italian students in two main areas, namely mathematics and language. Each test includes a set of multiple-choice items followed by open response questions. Students must conclude the tests in 45 to 90 minutes, depending on grade and subject.[3] The test was introduced in 2007 by law 176/2007, and it is administered yearly to second-, fifth-, eighth-, and tenth-grade students attending public or private schools.

We focus our analysis on second-grade primary school students (7-year-olds). Fifth graders (10-year-old) are also considered to offer some insights about the mechanism underlying our results.[4] Although each school knows the individual scores of its students, public data about individual performance on the INVALSI test are fully anonymous: students, classes, and schools cannot be identified. This makes it impossible to detect the grade composition of each class using the INVALSI data alone.

To overcome this limitation, we assembled a new data set that merges individual performance on the INVALSI test in the 2012/2013 SY with information included in two dif-

[3]More information about the INVALSI test is available at www.invalsi.it.

[4]From now on, unless otherwise specified, we will always refer to students enrolled in second grade.

ferent administrative archives: i) School Register data[5] provided by the Italian Ministry of Education (MIUR), which contain detailed information about each Italian primary school, including the number of multigrade classes[6]; and ii) the Municipality Register data produced by ISTAT, which include geographical and demographic information for each Italian municipality. We use data about municipalities to bridge information in the INVALSI data and in the School Register data, and we create specific algorithms to identify students attending a multigrade class.[7]

This procedure allows us to identify municipalities that host a single primary school, the name and the characteristics of this school, the educational achievements of its second-grade students, and the grade composition of their classes (single versus multigrade). As a result, our final data set (*full sample* hereafter) includes the entire population of Italian second-grade students attending a primary school located in municipalities hosting only one primary school. We end up with 4,295 primary schools out of 15,248 covered in the School Register data in the 2012/2013 SY, and about 92 thousand second-grade students out of the 500 thousand all over the country.[8]

In Italy, around 65 percent of municipalities with primary schools have no more than one such school, which reflects the fact that 53 percent of municipalities in Italy are rural (or inner areas) according to the classification provided by the Ministry of Economic Development, meaning that they are far from service provision centers (see Materiali UVAL, 2014). In these rural contexts, mostly represented by small municipalities (in our sample, the average population size for municipalities with a multigrade class is 1,029 inhabitants), multigrade classes are frequent.

[5]This data set is built on administrative data coming from the Ministero dell'Istruzione, dell'Università e della Ricerca (MIUR) *Rilevazione integrativa*.

[6]The analysis is based on the 2012/2013 SY as this is the only year in which the Ministry of Education provided the School Register data with the information needed to identify multigrade classes to us.

[7]Appendix A.1 below provides a detailed description of the data construction process.

[8]We drop from our analysis the two bordering regions of Valle d'Aosta and Trentino Alto Adige as in these areas, the administration of primary and secondary schools is assigned to the regional (or the provincial) authorities. As a consequence, the Ministry of Education does not collect registration information for these areas. Students from these two regions only account for 1.4 percent of the total sample size of INVALSI test takers.

On the one hand, our consideration of only municipalities that host no more than one primary school represents a potential data limitation. On the other hand, this limit allows us to keep under control the problem of nonrandom assignment of students into classes with different grade compositions. In fact, in municipalities with no more than one primary school, parental choice about their children's school enrollment is automatically ruled out unless parents decide to take them to a different municipality and bear commuting costs, which increase directly with the distance from the closest alternative solution (a variable we control for in our exercise below).

To improve comparability between the treatment and the control groups even further, we also define a restricted sample (*reduced sample* hereafter) including only schools with no more than one second-grade class.[9] In this second sample, parental choice is further restricted: parents cannot exercise any choice either about the school or the class for their children, unless again they bear the travel costs necessary to reach the closest alternative primary school.

Table 1 shows summary statistics for our samples. The average performance on the mathematics standardized test is around 19 points (out of 32, or about 59 percent of correct answers) while it is slightly higher (25 points out of 39, or about 64 percent of correct answers) for the language test. In the full sample, around 6 percent of second graders attend a multigrade class. When we look at the reduced sample, this percentage considerably increases to 16 percent. Class size is similar in the two samples with average values of 17.7–19.3 pupils per class. Children's characteristics are also similar when we compare the two samples. The average age is 7. The sample is balanced in terms of gender, as well as the percentage of children whose parents are migrants, around 10–12 percent of the two samples. In terms of socio-economic background, we consider three different levels of parental education: completed university, completed high school, and a residual category for all those holding below a secondary education diploma. Similar patterns emerge when fathers and

[9] The same sample selection criterion is adopted for the analysis of fifth graders.

mothers are compared. Most children in our samples (more than 70 percent) come from families in which parents have at most an upper secondary education. The percentage of university graduates is always lower than 10 percent, while 17–18 percent is the share of parents with an education below the high school level.

The bottom panel of Table 1 provides important information about the geographical characteristics of the schools. It is important to notice that our sample covers the entire Italian territory as all five macro-regions (NUTS 1) are represented. The Northwestern area is the most represented (43–46 percent), followed by the South (18–23 percent), the Northeast (13–17 percent), the central area (11–12 percent), and the Islands (8–9 percent). Finally, the reduced sample consistently differs from the full sample when it comes to resident population and altitude of the municipality. The difference originates from the definition of the full versus the reduced sample, making the latter more likely to include less populated and more peripheral areas.

4 The Identification Strategy

The identification of the causal impact of multigrading on child achievement is a challenging task because parental choice could drive the enrollment in multigrade classes. Although, as discussed in Section 2, this kind of concern should be relatively minor in the Italian context (because of the process of class formation that totally excludes parents), we fully address the endogeneity issue by implementing an instrumental variable (IV) identification strategy. Our IV strategy builds on the research design in Angrist and Lavy (1999), which is often referred to as the Maimonides' Rule.[10] In this work, the authors exploit class size cutoffs imposed by a rule in Israel to estimate the impact of class size on scholastic achievement. The same strategy is also used by Leuven and Rønning (2016), who exploit institutional features significantly affecting grade composition in Norway to specifically estimate the impact

[10]Estimation strategy inspired by Maimonides-style rules are common in the literature about class size and class composition. Some examples of works based on similar concepts are Hoxby (2000), Gary-Bobo and Mahjoub (2013), Bonesrønning (2003), Leuven et al. (2008), and Dobbelsteen et al. (2002).

of grade mixing on students' achievement.

Our identifying assumption is based on DPR 81/2009, a law that defines a set of rules based on exogenous cutoffs to establish whether a new single or multigrade class should be created. Specifically, we use predicted-by-the-law grade composition of classes to instrument the actual grade composition of classes. The Italian law is based on different thresholds defined in terms of the number of students of the same grade enrolled in a specific school. Single-grade classes should be comprised of a minimum of 15 and a maximum of 26 students; on the other hand, multigrade classes should be comprised of no fewer than 8 and no more than 18 students. In special cases—isolated villages, small islands, and areas characterized by the presence of linguistic minorities—deviation from the rules is possible: classes can be created with a lower number of students. However, this number cannot be lower than 10 students for single-grade classes.

It should be noted that although the law—in principle—prevents the creation of single-grade classes with fewer than 10 students, in practice they exist. In our sample, about 25 percent of students enrolled in schools with fewer than 10 second graders attend a single-grade class, while 75 percent attend a multigrade class.[11] Despite the flexibility to accommodate local requests in specific years, DPR 81/2009 identifies four different relevant intervals (based on the number of students enrolled in a specific grade) that cannot be modified by parents or school principals and that strongly affect the individual probability of being enrolled in a multigrade class. The first interval pertains to schools with fewer than 10 students in one grade. In this case, no single-grade class should be created, and all students should be assigned to a multigrade class. The second interval covers situations in which there are 10 to 14 students in one grade. In this interval, following school characteristics such as localization, both a single or a multigrade class could be created. The third interval covers school with between 15 and 26 students. In that case, the probability of being enrolled in a multigrade

[11] This evidence, on the one hand, makes essential the use of an instrumental variable approach to cope with possible endogeneity underlying similar cases. On the other hand, single-grade classes with fewer than 10 students are crucial for our analysis as they allow us to separately estimate the class size effect from the class composition effect.

class should be close to zero. The same applies for the last interval—situations with more than 26 students. The number of students is too high to create a multigrade class; therefore, according to the law, students should be assigned to more than one single-grade class. We exploit these four intervals as instruments to predict the actual class composition in terms of grade levels.

Class size is another important determinant of school performance that cannot be neglected in this framework as it is potentially correlated with the probability of attending multigrade classes, even though we observe both single and multigrade classes below the cutoff of 10 students. To consider the possible effect of class size on child achievement, we perform a dual IV analysis. First, we estimate a model including class size as a control variable. In addition, as class size might suffer from the same sources of endogeneity of multigrading, we replicate our model instrumenting both multigrading and class size. Given that the number of students enrolled in each grade is an important determinant of class size, we use the same set of instruments for multigrading and class size. The comparison between the two estimated models is an important robustness check of our findings.

The validity of our instrumental approach relies on different assumptions. To avoid violation of the exclusion restriction, we need our instrument to only affect students' test scores via grade composition (single versus multigrade class and class size). As already discussed, the exact number of students enrolled in a specific grade is unpredictable as in Italy each family is free to enroll children in every school nationwide, although students living in the catchment area of the school have a priority. Moreover, the enrollment procedure and its timing make it particularly difficult (if not impossible) for parents to form reliable expectations about the probability of their child ending up in a class with specific characteristics in terms of size and grade composition.

A second important assumption underlying the literature based on Maimonides-style IV approaches is the absence of ad hoc manipulation around cutoffs. As shown below, we do not find any evidence of such manipulation.

Under these assumptions we define the following reference model:

$$TestScore = \beta_1 + \beta_2 Multigrade + \beta_3 ClassSize + \boldsymbol{\beta}_4 \boldsymbol{X} + \beta_5 AreaFE + u \tag{1}$$

where $TestScore$ is the student's performance on the standardized national INVALSI test, $Multigrade$ is a dummy variable taking the value of one if the student is enrolled in a multigrade class, and $ClassSize$ represents the student's class size. The vector **X** contains contextual (observable) factors likely to affect test scores. Specifically, we control for child characteristics such as age, gender, and nationality (distinguishing among nationals of Italy, first-generation migrants, and second-generation migrants). Parental characteristics are another crucial set of determinants of test scores. We proxy parental background by including in the model both father's and mother's education (university graduated versus high school completed versus other) and profession.[12]

The vector **X** also includes information about the population, the altitude of the municipality hosting the school, and the minimum car travel time needed to reach the closest alternative primary school from the school each student actually attends. Travel time to the closest school is crucial as it underlies the presence of alternative school options. If alternative primary schools are available, parents who dislike multigrade classes and who have a high expected probability of their child ending up in such a class, might decide to enroll their child in the closest school offering single-grade classes. For this reason, we include in all our models this measure of travel distance; as a robustness check we also use this variable as an additional instrument for the individual probability of being enrolled in a multigrade class. Thanks to this strategy, we should deal with the possible residual endogeneity not corrected by our standard IV approach based on the definition of DPR 81/2009. Finally, to consider geographical differences across the country, we also include in our model a set of macro-region fixed effects that capture the average effect on test score for regions in the

[12] Unfortunately, the INVALSI data do not contain information about family income. However, educational level and profession for both parents represent good proxies.

Northwest, the Northeast, the central area, the South, and the Islands.[13]

As discussed, we estimate two different IV specifications defining two different sets of first stages. In the first specification, we instrument $Multigrade$ as in equation (2), while we use $ClassSize$ as a standard control variable:

$$Multigrade = \gamma_1 + \gamma_2 DPR81/2009 + \gamma_3 ClassSize + \boldsymbol{\gamma}_4 \boldsymbol{X} + \gamma_5 AreaFE + \epsilon \tag{2}$$

The instrument (DPR81/2009) is implemented through four variables based on intervals in the number of students enrolled in a specific grade.[14] The first variable labels schools with fewer than 10 students, the second between 11 and 14 students, the third between 15 and 26 students, and finally the fourth indicates schools with more than 26 students in second grade.

In the model including $ClassSize$ as an additional endogenous variable, we replicate the same first stage for multigrading (with the obvious exclusion of $ClassSize$ as a control variable) and then add a second first stage of the following form:

$$ClassSize = \delta_1 + \delta_2 DPR81/2009 + \boldsymbol{\delta}_4 \boldsymbol{X} + \delta_5 AreaFE + \varepsilon. \tag{3}$$

Because of possible serial correlation of the error term at the school level, all the models are estimated with standard errors clustered at school level.[15] We leave the discussion of instrument relevance in both specifications to the next section.

[13]We extensively test for different definitions of geographical areas in Section 5.3.

[14]Recall that in the main analysis we focus on second graders. We repeat the exercise for fifth graders to better understand the mechanism underlying our main results.

[15]Notice that we are considering municipalities with just one school in the full sample; hence, clustering at the school level is equivalent to clustering standard errors at the municipal level. When using the restricted sample, this is also equivalent to clustering at the class level as we are considering municipalities with one school hosting only one second-grade class.

5 The Effect of Multigrading on Child Achievement

5.1 First-Stage Estimates

Before showing first-stage estimates, we discuss a typical concern related to the adoption of Maimonides-style rules. Such an identification strategy conveys possible ad hoc manipulation around the cutoff to prevent the enforcement of specific class or grade compositions. We deal with this concern by comparing observable individual characteristics around the main cutoff (10 students). Table 2 reports the analysis of children's individual characteristics and family characteristics. We impose a 2-student-interval around the cutoff, comparing schools with 9 or 10 students with schools with 11 or 12 students. All the average values are remarkably similar around the cutoff. The p-values for the differences in means in column (4) confirm the lack of manipulation by school principals around the critical value of 10 students.[16] Similar conclusions arise when the analysis is applied to the three other relevant cutoffs identified by the law.

In Figure 1 we provide a graphical representation of the first stage estimates for second grade students.[17] We estimate the first stage as explained in Section 4: for each student, we compute the predicted probability of being assigned to a multigrade class based on the set of her individual characteristics. Panels (a) and (b) report the full and reduced samples, respectively. In both cases, one could notice different clouds of points. The concentration around different cutoffs shows the implementation of the law by school principals in primary schools. The predicted (by the law) class and grade composition represents an efficient and precise instrument for the actual class and grade composition.

Students in schools with at most 10 second graders have a very high predicted probability (80 percent) of ending up in a multigrade class. A second cloud of points is identified in the interval of 11-14 students. In this case, according to first-stage predictions, the probability of

[16] The fraction of females and the average age are significantly different, although point estimates are remarkably close.

[17] Here, we consider the reference model with only multigrade as an endogenous variable. Results do not change with the inclusion of class size as an additional endogenous variable.

being assigned to a multigrade class is close to 20 percent. On the contrary, the probability of being assigned to a multigrade class for other students is centered around zero for both the full and the reduced samples.[18]

Table 3 shows the first-stage estimates for second grade students. Columns (1) and (4) report results for multigrading as the only endogenous variable, while columns (2–3) and (5–6) also treat class size as endogenous. Models in columns (1) to (3) exploit the full sample; the analysis shown in columns (4) to (6) is based on the reduced sample. All the tests for under and weak identification suggest that the first stage is very precise, and the instruments are extremely relevant. We start by analyzing the model considering only multigrading as an endogenous regressor (columns 1 and 4). With respect to the omitted category (schools with more than 26 second graders), students in schools with at most 10 second graders are highly likely to be assigned to a multigrade class. The coefficient is 0.80 in the full sample (column 1), and around 1 for the reduced sample (column 4). The presence of at most 10 second graders enrolled in a specific school increases by 80–100 percent the individual probability of being assigned to a multigrade class. This result is hardly surprising as the law forces the adoption of multigrade classes for these specific cases. The coefficient remains significant, but with a lower magnitude (0.14–0.41), for schools with 11–14 second-grade students. Schools with 15–26 students display close-to-zero effects in the full sample. The coefficient in the reduced sample is 0.18.

Results for the probability of being enrolled in a multigrade class do not qualitatively change when class size is treated as a second endogenous variable (columns 2 and 5), and DPR 81/2009 stands out as a precise predictor for class size (columns 3 and 6). Higher numbers of enrolled second graders positively affect observed class size. As an example, in the full (reduced) sample, attendance at a school with at most 10 second-grade students explains an average decrease of 5 (13) pupils per class with respect to a school with more than 26 students.

[18]The same graphical analysis for fifth grade students is reported in Figure A.1.

As for the role of other variables, the analysis of the first stage unveils the role of parental education in shaping individual probability of attending a multigrade class. Students of parents reporting at most a high school diploma display a zero and statistically insignificant increase in the likelihood of being enrolled in a multigrade class as opposed to students of parents with a university degree. This finding confirms that parents are unlikely to understand the individual probability of their children's being assigned to a multigrade class or—alternatively—that their background does not systematically shape their preferences on this matter.

5.2 Second-Stage Estimates

Table 4 shows second-stage estimates of the model in equation (1). We estimate six different specifications based on different samples and sets of endogenous variables. In columns (1–3) we use the full sample, while in columns (4–6) we restrict the focus to the reduced sample. For each sample, we estimate the reference OLS model (columns 1 and 4), the model with class size as a control variable (columns 2 and 5), and the model with class size as an endogenous variable (columns 3 and 6).

We measure child cognitive achievements by combining the math and language INVALSI standardized test scores.[19] After normalizing both test scores (with a mean of zero and a standard deviation of one), we create a combined score in math and reading, taking the average of the normalized reading and math scores. We then normalized the combined score.

The full sample analysis (columns 1–3) displays a strong and positive impact of multigrading on cognitive achievement. In the OLS framework, attendance in a multigrade class increases, by as much as 9 percent of a standard deviation, the combined math-language test score. The effect is 15–18 percent of a standard deviation when the IV strategy is implemented. The reduced sample analysis highlights similar findings. The OLS estimate is 10 percent of a standard deviation, and IV estimates are 19–20 percent of a standard deviation.

[19] The procedure is similar to the one used in Dahl and Lochner (2012) and Agostinelli and Sorrenti (2018).

Class size plays a significant role in affecting child achievement, although it should be noted that point estimates for the effect of multigrading are almost unaffected by the inclusion of class size as a pure control or as an endogenous variable. Quantitatively, a one-student-per-class increase explains an average decrease in individual performance of around 1 percent of a standard deviation.

Results in Table 4 require some additional discussion. First, it is important to note that although OLS and IV provide the same qualitative conclusion, the IV coefficients are higher in magnitude. This difference is driven by many different factors, such as omitted variable bias in OLS estimates and measurement error due to possible (reporting) errors in the administrative data we use to identify multigrade classes.

A second important aspect is the stability of results when different specifications, based on different samples, are estimated. Coefficients never significantly change when the full and the reduced samples are compared. The same consideration applies to class size coefficients. Including class size as a pure control variable or considering it as an endogenous variable leaves almost unaffected results on the impact of multigrading on child achievements.

Third, we further discuss the possible existence of parents' preference for single versus multigrade classes. Such preference is a potential additional source of endogeneity underlying the individual enrollment in a multigrade class. First-stage evidence (see Section 5.1) regarding the role of parental education in shaping multigrade class attendance signals that this is unlikely to be a threat to the reliability of our findings. However, to be even more cautious, we estimate in Table 5 an additional IV specification in which travel time to the closest school is also used as an instrument. Assuming that parental preferences about grade composition play a role in choosing a school for their children, we have to consider that these preferences are constrained by the time needed to reach the closest alternative school. This constraint is likely to be even more binding in the case of second-grade students, as curricula are uniform across different schools and school principals have little room for differentiating the quality of their educational services.

First-stage estimates suggest that travel time to the closest school plays a modest role in determining the individual probability of being enrolled in a multigrade class.[20] An additional 1-minute distance causes an increase in the probability of enrollment in a multigrade class of 0.01 percent. The coefficient is statistically significant at the 5 percent level.[21] No effect of travel time on class size is detected.

Table 5 reports second-stage estimates. Results are unaffected by the inclusion of travel time to the closest school as an additional instrument. As in our baseline analysis, the effect of multigrading on child cognitive achievements ranges between 15 and 19 percent of a standard deviation. This additional evidence reinforces the idea that the implemented baseline empirical strategy copes with the main endogeneity issues underlying individual attendance of a multigrade class.

5.3 Sensitivity Tests

In this section we test the sensitivity of our results to some modeling choices, showing that our findings are unaffected by these choices. First, in the baseline analysis, we have considered the combined math-language test score as the main outcome of interest. We focus here on its two components separately. Table 6 shows the effect of multigrading classes on each single test score, namely mathematics and language. The effect is positive and statistically significant for both items. This guarantees that the overall effect shown in our baseline analysis is not exclusively driven by one single subject. The effect on mathematics scores seems slightly higher in magnitude (16–20 percent of a standard deviation) compared to the effect on language scores (11–15 percent of a standard deviation), although the difference between the two is not statistically significant.

Table 7 shows a further set of sensitivity tests based on different specifications of the baseline reference model. In the first test (Panel a) we augment the model with a control for

[20]For the sake of brevity, we comment on first-stage estimates without reporting the full set of estimates. However, all the results are available upon request.

[21]Results are stable across different samples and across the standard dual approach for the analysis of the role of class size.

school size. The aim is to check whether our results are affected by the inclusion of school size once we control for the class-size effect, as a potential determinant of children's test scores. The results are unaffected by the inclusion of school size as an additional control variable.

In Panels (b) and (c) we investigate the possible geographical connotation of the multi-grading effect. This analysis is important per se, as it allows to infer possible heterogeneity at the local level. At the same time, it allows us to deal with some of the concerns related to possible bias induced by cheating and opportunistic behavior on the INVALSI test. We will discuss this point in detail below. In our main analysis, we use the five macro-regions (NUTS 1) to capture macro-regional fixed effects. Here we estimate two alternative models by considering regional fixed effects (NUTS 2, Panel b), and provincial fixed effects (NUTS 3, Panel c). Results remain unchanged in both specifications, suggesting that the choice of geographical level of aggregation is not affecting the size and significance of our findings.

Finally, in Panel (d) we tackle one of the limitations of the INVALSI data: missing information about parents, or about some of their features (such as education or job). Although in our baseline analysis we introduced residual groups for students with missing information on parents here we restrict our sample to include only students for whom information about both parents' profession and educational level are available.[22] Results are the same as in the baseline analysis.

Finally, we discuss concerns related to possible opportunistic behavior on standardized tests. The use of scores on standardized tests to assess individuals' skills is common in social sciences (for example economics, sociology, psychology, etc.). However, given that standardized tests are useful tools to compare different schools, classes, and teachers, this produces potential incentives for opportunistic behavior by principals, teachers, and even students. For this reason, many scholars advocate against the reliability of these tests by providing growing evidence of cheating behavior and score manipulation. For instance, Jacob and

[22] We keep these observations in the baseline model as we want to also consider single-head households in our analysis.

Levitt (2003) estimate that, in Chicago public schools, serious cases of cheating by teachers or administrators occur in at least 4–5 percent of elementary school classrooms. Similarly, a well-established systematic cheating practice perpetrated by teachers was discovered in 2011 in the city of Atlanta (Severson, 2011). Outside the United States, the debate about test score reliability has been raised in many countries such as the UK, Israel, France, and Sweden (e.g. Diamond and Persson, 2016).

The structure of the INVALSI test might generate incentives for teachers' or students' opportunistic behavior aimed at score manipulation. In particular, teachers might decide to help their student by suggesting correct answers, fixing wrong ones, or filling in missing answers to improve class performance. Bertoni et al. (2013) provide the first empirical evidence of the possible existence of cheating behavior on the INVALSI test. They show that schools whose test is administered by an external examiner perform worse than schools whose test is administered by resident teachers or professors. According to Lucifora and Tonello (2015), cheating behavior mainly occurs when teachers shirk or decrease monitoring efforts. In analyzing class size and score manipulation in Southern regions, Angrist et al. (2017) find that cheating largely reflects teacher behavior, motivated by moral hazard in grading effort.

Although it is impossible to be conclusive in identifying cheating, here we show evidence about its potential impact on our estimates. According to Angrist et al. (2017), roughly 5 percent of Italian scores are biased because of cheating. As the test is not thought to evaluate single schools but rather to provide a map of the efficiency of the Italian schooling system, INVALSI provides a deterministic massive correction measure to address opportunistic behavior. However, this correction is based on a fixed predetermined rule (considering, among other things, intra-class variance in scores) that (as confirmed by INVALSI) is inappropriate for multigrade classes analyzed here. Schools with multigrade classes, as well as a considerable fraction of schools in our sample, are almost by definition small schools.[23]

[23]Remember that our sample merely includes institutions that are the only school in their municipality. The restricted sample includes only schools with no more than one second grade class.

They are characterized by very low numbers of enrolled students (and consequently small class sizes), which makes the use of deterministic corrections such as the one operated by INVALSI ineffective.

Several factors suggest that cheating should not be a major concern for our analysis. First, the existence of cheating in our setting would imply that our outcome of interest is the real score at the test plus some noise. On the one hand, if noise is stochastic, this would only affect our estimates by lowering precision, and all the coefficients would remain consistently estimated. On the other hand, if noise is correlated with our variable of interest (being enrolled in a multigrade class) the coefficients estimates would potentially be biased. As also confirmed by our discussions with administrators, principals, primary school teachers, and members of INVALSI, it is difficult to think that the probability of observing opportunistic and cheating behavior directly depends on considering a single versus a multigrade class. Other elements suggested by the literature, such as teachers' unobserved characteristics, should be considered as the main determinants of possible cheating (Angrist et al., 2017). This anecdotal evidence is also confirmed by intraschool variability in cheating patterns.

We empirically deal with possible cheating-induced bias in our estimates with the analysis of geographical patterns underlying our baseline results. As shown by Bertoni et al. (2013), Angrist et al. (2017), and other qualitative studies, cheating behavior is a major concern in Southern Italy and much less so in Northern regions. According to the score manipulation index elaborated by Angrist et al. (2017), cheating only accounts for 2 percent of scores in the North and central area of Italy; this percentage is even lower for Northern regions only. Figure 2 (right panel) shows the geographical distribution at provincial level of score manipulation in Angrist et al. (2017). Almost all the provinces in the Northern part of the country are characterized by zero score manipulation. Ferrer-Esteban (2012) computes a similar measure for cheating behavior by focusing on the exact repetition at the class level of the same sequence of answers (left panel of Figure 2).

With this evidence in mind, in Table 8, we replicate our baseline analysis focusing on regions in Northern Italy.[24] Despite the reduced sample size, results for the Northern region are similar to the ones we obtain for the whole country. OLS estimates do not show any relevant difference, with coefficient estimates (8 percent of a standard deviation) almost unchanged. The same happens for IV estimates: the effect of multigrading is 10–15 percent of standard deviation in the full sample, while it appears as slightly smaller (11–13 percent of a standard deviation) in the reduced sample. The coefficients of our relevant covariates in the sample of Northern regions are never statistically different from those in the sample including the whole set of Italian regions.[25]

The discussed anecdotal and empirical evidence suggests that the possible bias induced by cheating and opportunistic behavior might only marginally affect our results, which are consistent with the findings in Leuven and Rønning (2016) for the case of 15-year-old students.

5.4 Heterogeneous Effects of Multigrading

Our baseline analysis shows that classes mixing pupils of different ages are beneficial in terms of cognitive development for 7-year-old children. Is the benefit the same for all children? We propose here a simple heterogeneity analysis based on two important dimensions: gender and family background.[26] Undeniably, the analysis of heterogeneous effects is always complicated in an IV setting. Instruments usually affect locally a fraction of the population, making it very difficult to compare different subpopulations. In our setting, the use of a law based on simple numerical rules to assign students to a single or to a multigrade class makes the analysis easier. Indeed, the instrument is also extremely powerful and relevant when using subsamples.

Table 9 (columns 1–2) shows the analysis by gender. Scholastic performance is typically

[24] A first set of sensitivity tests based on different definitions of geographical variables is discussed above.

[25] We perform the same analysis for the case of fifth graders. Results are reported in Table A.5. The same considerations for the case of second graders also apply in this case. Any significant difference is detected when the analysis is carried out by only considering Northern Italian regions.

[26] In Section 6 we focus on the effect induced by age composition of the class to understand the differential effect of studying with younger or more mature peers.

different when males and females are compared, with males usually underperforming in most subjects except mathematics. Results by gender suggest that, although differences across genders are not striking, females seem to benefit more from multigrading than males. In our full sample, the coefficient for multigrading increases by 50 percent (from 12 to 18 percent of a standard deviation) when females are compared to males. A similar result is found in the reduced sample, with the coefficient switching from 16 to 23 percent of a standard deviation. In general, both genders seem to benefit from multigrading, with girls obtaining higher benefits compared to boys.

We then study the heterogeneity due to different parental backgrounds. We divide children in two groups according to their parents' education.[27] The low-background group (No one with university) includes children whose parents do not hold a university degree. The high-background group (One with university) children with at least one parent holding a university degree.

Table 9 (columns 3–4) reports the results. Multigrading positively shapes child's cognitive achievements for both parental backgrounds. However, the effect seems to be mainly driven by children from the lower parental background. As an example, in the reduced sample, the coefficient for the lower background is almost twice as large as the one for the higher background (21 versus 11 percent of a standard deviation). Moreover, the coefficient for children with at least one parent with a university degree is never statistically significant.[28]

The analysis of parental background highlights that children from less stimulating home environments obtain the highest benefit from attendance in a multigrade class. This result identifies grade composition as a potential tool to mitigate the long-term effects of pupils' lower socio-economic backgrounds. A class environment consisting of peers of different ages (in particular, older peers) might act as an important additional input in the child development production function, partially compensating for the negative impact of

[27]Parental education is widely acknowledged as a good proxy for family socio-economic background.

[28]It should be noted that in this specific case, the sample size for the higher parental background is considerably smaller than for the lower parental background. However, the instrument is relevant and strong in this subsample as well, making the concern of different sample sizes across the two groups less relevant.

low socio-economic conditions.

6 Younger is Better? Grade Composition and the Effect of Multigrading

We investigate the mechanism underlying the positive average effect of multigrading on cognitive achievements. A dual analysis is presented. First, we replicate results by splitting the sample of second-grade students who attended a multigrade class into three different groups: students with only older peers in the class, students with only younger peers in the class, and the mixed case comprised of students with both younger and older peers in the class.[29]

Table 10 illustrates results for second graders by grade composition of their class. The full sample analysis is performed in columns (1–3), while the analysis for the reduced sample is shown in columns (4–6). Because it is difficult to identify proper instruments for both the probability of ending up in a multigrade class and the class composition in terms of grades, we only report OLS estimates for the two samples. Indeed, there are no rules guiding students' assignments to multigrade classes with older, younger, or both types of peers; the decision rests solely with the school principals based on the number of pupils in each year for each grade. However, the similarity between the (qualitative) results of OLS and IV models in the baseline analysis supports the idea that OLS models produce reliable estimates for the effect of interest.

Results highlight that the overall positive effect of multigrading is mainly driven by students sharing their (multigrade) class with more mature peers. Attending a multigrade class with more mature peers (columns 1 and 4) explains a 16 percent of a standard deviation

[29] Older peers are children in higher grades (third, fourth, and fifth grades) attending the same multigrade class of second-grade students. Symmetrically, younger peers are only first graders attending the same multigrade class of second-grade students. Note that we do not have any test scores for first, third, and fourth graders as they do not take the INVALSI test. We will exploit the fifth graders' scores below. See Appendix A.1, part (b) for details about the process we use to identify such students.

increase in test scores. On the contrary, attending a multigrade class with younger peers (columns 2 and 5) produces a positive although statistically insignificant effect on test scores. As additional evidence about the (relative) importance of attending classes with older peers, the effect of multigrading is also positive and statistically significant (19–23 percent of a standard deviation) when it comes to mixed classes where second grade students are in contact with both younger and older peers.

According to this first analysis, multigrading is particularly beneficial when a child shares the class with more mature peers. This framework is likely to inspire and foster child interactions with, and imitation of, her more mature peers. At the same time, sharing the classroom with younger peers appears nondetrimental for child development at this particular stage of individual growth.

Second, we replicate the baseline analysis by considering fifth-grade students at primary schools. These students constitute the older cohort in multigrade classes as the fifth grade is the last grade of primary school. With this analysis we investigate students' scores that are driven only by their interaction with younger peers. Table 11 displays the results of the standard IV analysis.[30] The coefficient for multigrading is never statistically significant and has point estimates remarkably close to zero. Multigrading does not affect in any way the performance on standardized tests by fifth graders. This result is important as it shows that multigrading is never detrimental for child achievement when the primary school cycle of education is analyzed.[31] Additionally, the effect of class size (-1 percent of a standard deviation) is similar to the one obtained for second-grade students.

To sum up, the analyses of second-grade class composition and fifth-graders' performance seem to reach the same general conclusion. When a child is particularly young, 6 to 10 years old, attending multigrade classes is in general nondetrimental for her cognitive development.

[30]Descriptive statistics, first-stage estimates, and graphical representation for the case of fifth graders are reported in Tables A.1, A.2, and Figure A.1 in Appendix A.2.

[31]In Table A.3 we replicate the model based on the time distance to the closest alternative school as an additional instrument to cope with possible residual endogeneity underlying parental preferences. In Table A.4 we decompose the general effect by considering the math and the language test scores in isolation. Results are always similar to the ones of the baseline models for fifth-grade students in Table 11.

Moreover, when attending classes with more mature peers, multigrading becomes beneficial and it explains a considerable and sizable increase in test scores.

7 Conclusion

The development of cognitive and non-cognitive skills in early childhood is recognized as a strong predictor of future success in academics as well as in life. For this reason, pedagogical practices that increase the abilities of young children represent powerful tools for improving individual well-being and reducing the chance of failure in the future.

Multigrading—placing kids of different ages in the same classroom—is a common educational practice in both developing and developed countries. In both instances, multigrading has often been adopted for budgetary reasons: although in developing countries it represents a widespread practice, in developed countries its use is generally confined to rural areas that are subject to population decline and where few children actually live. Nonetheless, over the last few years, multigrading has been adopted in several developed countries for reasons that go beyond budget constraints and—quite the opposite—are related to educational and pedagogical concerns. The supporters of this method emphasize its positive effects, including the benefits of a personalized approach to education, given that children of the same age can learn at different speeds, as well as to positive peer effects, with younger children imitating their older peers, and the latter becoming more prone to responsible behavior. Although quite a common practice, the effects of multigrading on child achievement has rarely been carefully investigated because of possible endogeneity concerns.

In this paper we aim to understand the effect of multigrading on children attending the second year of primary school in Italy. We do so by supplementing information on standardized test scores provided in the INVALSI data with information on schools and multigrade classes. To address endogeneity concerns and allow a proper causal inference, we exploit a national regulation by considering the number of second graders to determine whether to form a single or a multigrade class.

Our results—robust to different sensitivity tests—strongly suggest that multigrade teaching positively affects achievement, and this positive effect is stronger for children from low socio-economic backgrounds. The mechanism behind our results seems to be driven by the presence of more mature peers in the classroom, which are likely to substitute poorly educated parents in providing educational support to young children.

This work suggests at least two main relevant policy implications. Multigrading is not detrimental to child cognitive achievement; it could also represent a method of hindering poverty by giving advantage to children coming from disadvantaged cultural and economic backgrounds. In fact, older peers could represent a potential substitute for parental involvement in the education of younger children. Moreover, multigrading is quite common in rural, sometimes remote, areas still common in Europe. These areas are generally characterized by low population density, deprivation, and abandonment by younger generations, which make their situation even worse. Schools are likely to be the only institution that has the potential to revitalize these areas.

References

Agostinelli, F. and G. Sorrenti (2018). Money vs. Time: Family Income, Maternal Labor Supply, and Child Achievement. University of Zurich Working Paper No. 273.

Angrist, J. D., E. Battistin, and D. Vuri (2017). In a Small Moment: Class Size and Moral Hazard in the Italian Mezzogiorno. *American Economic Journal: Applied Economics 9*(4), 216–49.

Angrist, J. D. and V. Lavy (1999). Using Maimonides' Rule to Estimate the Effect of Class Size on Scholastic Achievement. *The Quarterly Journal of Economics 114*(2), 533–575.

Angrist, J. D., V. Lavy, and J. L.-L. A. Shany (2017). Maimonides Rule Redux. NBER Working Paper No. 23486.

Bertoni, M., G. Brunello, and L. Rocco (2013). When the Cat is Near, the Mice Won't Play: The Effect of External Examiners in Italian Schools. *Journal of Public Economics 104*, 65–77.

Bonesrønning, H. (2003). Class Size Effects on Student Achievement in Norway: Patterns and Explanations. *Southern Economic Journal 69*(4), 952–965.

Checchi, D. and M. De Paola (2017). The Effect of Multigrade Classes on Cognitive and Non-Cognitive Skills: Causal Evidence Exploiting Minimum Class Size Rules in Italy. IZA Working Paper no. 11211.

Cunha, F. and J. J. Heckman (2008). Formulating, Identifying and Estimating the Technology of Cognitive and Noncognitive Skill Formation. *Journal of Human Resources 43*(4), 738–782.

Cunha, F. and J. J. Heckman (2010). Investing in Our Young People. In A. J. Reynolds, A. Rolnick, M. M. Englund, and J. Temple (Eds.), *Cost-effective Early Childhood Programs in the First Decade: A Human Capital Integration*. Cambridge University Press.

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

Dahl, G. B. and L. Lochner (2012). The Impact of Family Income on Child Achievement: Evidence from the Earned Income Tax Credit. *American Economic Review 102*(5), 1927–1956.

Diamond, R. and P. Persson (2016). The Long-term Consequences of Teacher Discretion in Grading of High-stakes Tests. NBER Working Paper No. 22207.

Dobbelsteen, S., J. Levin, and H. Oosterbeek (2002). The Causal Effect of Class Size on Scholastic Achievement: Distinguishing the Pure Class Size Effect from the Effect of Changes in Class Composition. *Oxford Bulletin of Economics and Statistics 64*(1), 17–38.

Ferrer-Esteban, G. (2012). Cheating to the Test in the Italian Standardized Assessment System: Rationale and Incentives. Working Paper Fondazione Giovanni Agnelli no. 50.

Gary-Bobo, R. J. and M. B. Mahjoub (2013). Estimation of Class-Size Effects, Using "Maimonides' Rule" and Other Instruments: The Case of French Junior High Schools. *Annals of Economics and Statistics* (111/112), 193–225.

Hargreaves, E., C. Montero, N. Chau, M. Sibli, and T. Thanh (2001). Multigrade Teaching in Peru, Sri Lanka and Vietnam: An Overview. *International Journal of Educational Development 21*(6), 499–520.

Heckman, J. J., J. Stixrud, and S. Urzua (2006). The Effects of Cognitive and Noncognitive Abilities on Labor Market Outcomes and Social Behavior. *Journal of Labor Economics 24*(3), 411–482.

Hoxby, C. M. (2000). The Effects of Class Size on Student Achievement: New Evidence from Population Variation. *The Quarterly Journal of Economics 115*(4), 1239–1285.

Jacob, B. A. and S. D. Levitt (2003). Rotten Apples: An Investigation of the Prevalence and Predictors of Teacher Cheating. *The Quarterly Journal of Economics 118*(3), 843–877.

Leuven, E., H. Oosterbeek, and M. Rønning (2008). Quasi-experimental Estimates of the Effect of Class Size on Achievement in Norway. *Scandinavian Journal of Economics 110*(4), 663–693.

Leuven, E. and M. Rønning (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.

Lucifora, C. and M. Tonello (2015). Cheating and Social Interactions. Evidence from a Randomized Experiment in a National Evaluation Program. *Journal of Economic Behavior & Organization 115*, 45–66.

Materiali UVAL (2014). A Strategy for Inner Areas in Italy: Definition, Objectives, Tools and Governance. In F. Barca, P. Casavola, and S. Lucatelli (Eds.), *Analisi e studi, Documenti, Metodi*. Materiali Uval.

Mulkeen, A. and C. Higgins (2009). Multigrade Teaching in Sub-Saharan Africa: Lessons from Uganda, Senegal, and the Gambia. World Bank Working Paper no. 173.

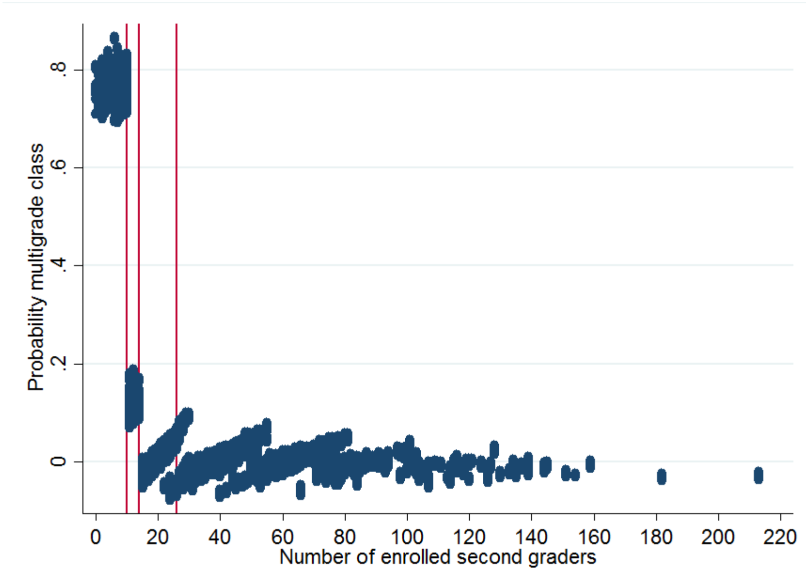
Severson, K. (2011). Systematic Cheating Is Found in Atlanta's School System. New York Times, July 5, 2011.

Todd, P. E. and K. I. Wolpin (2007). The Production of Cognitive Achievement in Children: Home, School, and Racial Test Score Gaps. *Journal of Human Capital 1*(1), 91–136.

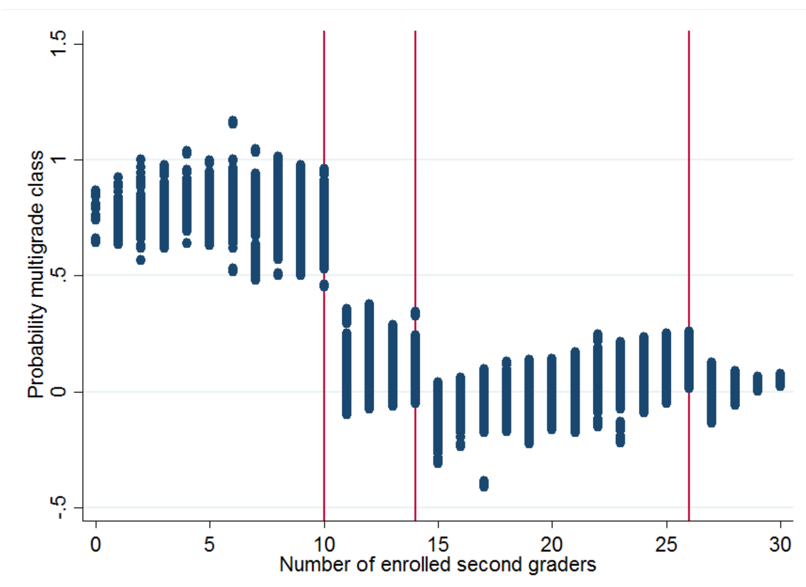
Figures and Tables

Figure 1: Number of Second Graders and Individual Probability of Multigrading

(a) *Full sample*



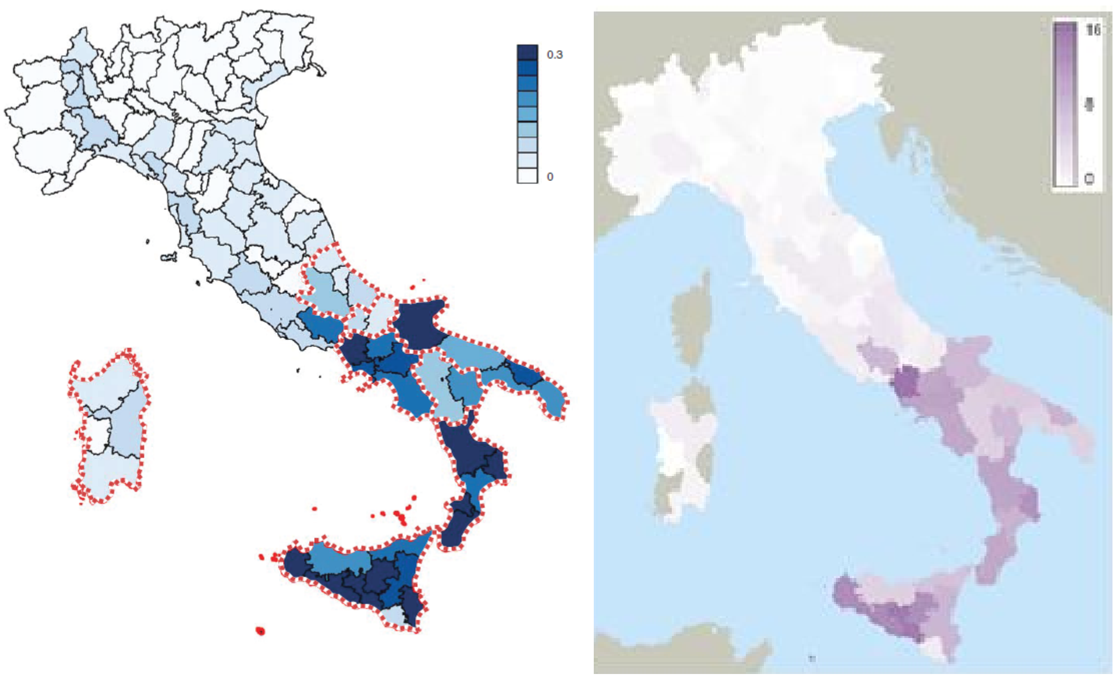
(b) *Reduced sample*



Notes: This figure shows the predicted individual probability of ending up in a multigrade class for second-grade students as a function of the number of second-grade students enrolled in a school. Panel (a) refers to the full sample, Panel (b) is based on the reduced sample. The predicted individual probability of attending a multigrade class (y-axis) is obtained through first-stage estimates in Table 3, columns (1) and (4). Refer to the text and to Table 3 for further details about the empirical model underlying this figure.

Figure 2: Territorial Distribution of Cheating Behavior

(a) *Score manipulation rates by province* (b) *Cheating behavior*



Notes: This figure shows the geographical distribution of cheating behavior according to two different sources. Panel (a) refers to the work by Angrist et al. (2017) and it is based on (i) implausible score levels, (ii) the within-class average and standard deviation of test scores, (iii) the number of missing items, and (iv) a Herfindahl index of the share of students with similar response patterns. Panel (b) refers to the work by Ferrer-Esteban (2012) and is based on the analysis of the sequence of identical answers at the class level as a signal for possible cheating behavior.

Table 1: Summary Statistics

	Full sample		Reduced sample	
	Mean (1)	St.Dev. (2)	Mean (3)	St.Dev. (4)
Math	18.95	6.74	19.47	6.80
Language	24.84	6.73	25.20	6.81
Multigrade	0.06	0.23	0.16	0.37
Class size	19.35	4.09	17.67	4.86
Age	6.97	0.27	6.96	0.28
Female	0.49	0.50	0.49	0.50
Italian	0.88	0.32	0.90	0.30
Migrant 1st gen.	0.03	0.17	0.03	0.17
Migrant 2nd gen.	0.09	0.28	0.07	0.25
Father university	0.07	0.26	0.06	0.24
Father high school	0.74	0.44	0.77	0.42
Father other	0.18	0.39	0.17	0.38
Mother university	0.10	0.30	0.09	0.29
Mother high school	0.72	0.45	0.74	0.44
Mother other	0.17	0.38	0.17	0.37
Northwest	0.46	0.50	0.43	0.49
Northeast	0.17	0.38	0.13	0.34
Central area	0.11	0.32	0.12	0.33
South	0.18	0.38	0.23	0.42
Islands	0.08	0.27	0.09	0.29
Time distance (min.)	5.49	3.30	6.18	3.67
Population (2011)	4,671	3,097	1,959	831
Altitude	261	222	360	256
Observations	92,504		32,659	

Summary statistics for the samples analyzed in this work. Columns (1) and (2) refer to the full sample; columns (3) and (4) refer to the reduced sample.

Table 2: Balancing Test Around the 10-Student Cutoff

	Below Cutoff (BC)	Above Cutoff (AC)	BC-AC	P-value (BC-AV)
	(1)	(2)	(3)	(4)
Age	6.95	6.96	-0.02	0.06
	(0.01)	(0.01)	(0.01)	
Female	0.48	0.50	-0.03	0.06
	(0.01)	(0.01)	(0.01)	
Italian	0.90	0.90	0.00	0.69
	(0.01)	(0.01)	(0.01)	
Migrant 1st gen.	0.04	0.04	0.00	0.72
	(0.00)	(0.00)	(0.00)	
Migrant 2nd gen.	0.06	0.06	-0.01	0.44
	(0.00)	(0.00)	(0.01)	
Father university	0.06	0.05	0.00	0.49
	(0.00)	(0.00)	(0.01)	
Father high school	0.76	0.77	-0.01	0.38
	(0.01)	(0.01)	(0.01)	
Father other	0.18	0.17	0.01	0.58
	(0.01)	(0.01)	(0.01)	
Mother university	0.08	0.09	-0.00	0.60
	(0.01)	(0.01)	(0.01)	
Mother high school	0.75	0.75	-0.00	0.87
	(0.01)	(0.01)	(0.01)	
Mother other	0.08	0.09	-0.00	0.60
	(0.01)	(0.01)	(0.00)	
Interval around the cutoff (in nr. of enrolled students)	[9,10]	[11,12]		

Comparison of the population just below (column 1) and just above (column 2) the cutoff of 10 second-grade enrolled students. Intervals around the cutoff are made by 2 students (9–10 students vs. 11–12 students). The difference in means and the P-value for difference in means are reported in columns (3) and (4), respectively.

Table 3: First-Stage Estimates

	Full sample			Reduced sample		
	Model (1)	Model (2)		Model (1)	Model (2)	
	Multigrade	Multigrade	Class size	Multigrade	Multigrade	Class size
	(1)	(2)	(3)	(4)	(5)	(6)
$2^{nd}Graders \leq 10$	0.80***	0.75***	-5.31***	1.04***	0.69***	-12.69***
	(0.01)	(0.02)	(0.20)	(0.02)	(0.02)	(0.27)
$11 \leq 2^{nd}Graders \leq 14$	0.14***	0.10***	-4.99***	0.41***	0.06***	-12.74***
	(0.02)	(0.01)	(0.18)	(0.03)	(0.02)	(0.22)
$15 \leq 2^{nd}Graders \leq 26$	-0.01***	-0.00	1.17***	0.18***	-0.01	-6.83***
	(0.00)	(0.00)	(0.18)	(0.02)	(0.01)	(0.20)
Class size	0.01***			0.03***		
	(0.00)			(0.00)		
Father high school	0.00	0.00	0.03	0.01	0.01	0.07
	(0.00)	(0.00)	(0.05)	(0.01)	(0.01)	(0.07)
Mother high school	0.00	0.00	0.02	-0.00	0.00	0.08
	(0.00)	(0.00)	(0.04)	(0.00)	(0.00)	(0.06)
Instrumented variable(s)	Multigrade	Multigrade+Class size		Multigrade	Multigrade+Class size	
SW Chi-sq. (UId)	> 100	> 100	> 100	> 100	> 100	> 100
P-value	0.00	0.00	0.00	0.00	0.00	0.00
SW F (WId)	> 100	> 100	> 100	> 100	> 100	> 100
P-value	0.00	0.00	0.00	0.00	0.00	0.00
KP (WId)	> 100	> 100	> 100	> 100	> 100	> 100
Observations	92,504	92,504	92,504	32,659	32,659	32,659

First-stage estimates. Dependent variable: Being enrolled in a multigrade class (columns 1,2,4, and 5), class size (columns 3 and 6). The reference category for the number of second graders is the class $2^{nd}Graders > 26$. The reference category for father's and mother's education is completed university. All models include controls for child's gender, age, nationality, and father's and mother's profession. All models also include variables for altitude and population of the municipality, geographical macro-area, and road distance in time to the closest alternative school. Standard errors are clustered at the school level and reported in brackets. *, **, *** indicate statistical significance at the 10%, 5%, and 1% levels, respectively.

Table 4: Multigrading and Child Achievement

	Combined Math-Language					
	OLS (1)	IV (2)	IV (3)	OLS (4)	IV (5)	IV (6)
Multigrade	0.09***	0.15***	0.18***	0.10***	0.20***	0.19***
	(0.03)	(0.04)	(0.06)	(0.03)	(0.05)	(0.06)
Class size	-0.01***	-0.01***	-0.01	-0.01***	-0.01***	-0.01***
	(0.00)	(0.00)	(0.01)	(0.00)	(0.00)	(0.00)
Instrumented variable(s)		Multigrade	Multigrade, Class size		Multigrade	Multigrade, Class size
Sample	Full	Full	Full	Reduced	Reduced	Reduced
Observations	92,504	92,504	92,504	32,659	32,659	32,659

OLS and IV estimates of the effect of multigrading on a child's test score. Dependent variable: Combined Math-Language test score. All models include controls for child's gender, age, nationality, father's and mother's educational level, and father's and mother's profession. All models also include variables for altitude and population of the municipality, geographical macro-area, and road distance in time to the closest alternative school. Standard errors are clustered at the school level and reported in brackets. *, **, *** indicate statistical significance at the 10%, 5%, and 1% levels, respectively.

Table 5: Distance to the Closest School as Instrument for Parents' Preferences

	Combined Math-Language			
	IV (1)	IV (2)	IV (3)	IV (4)
Multigrade	0.15***	0.18***	0.19***	0.18***
	(0.04)	(0.06)	(0.05)	(0.06)
Class size	-0.01***	-0.01	-0.01***	-0.01***
	(0.00)	(0.01)	(0.00)	(0.00)
Instrumented variable(s)	Multigrade	Multigrade, Class size	Multigrade	Multigrade, Class size
Sample	Full	Full	Reduced	Reduced
Observations	92,504	92,504	32,659	32,659

IV estimates of the effect of multigrading on a child's test score. Dependent variable: Combined Math-Language test score. Road distance in time to the closest alternative school is used as additional instrument for being enrolled in a multigrade class and class size. All models include controls for child's gender, age, nationality, father's and mother's educational level, and father's and mother's profession. All models also include variables for altitude and population of the municipality, geographical macro-area. Standard errors are clustered at the school level and reported in brackets. *, **, *** indicate statistical significance at the 10%, 5%, and 1% levels, respectively.

Table 6: Math and Language Test Scores

	IV (1)	IV (2)	IV (3)	IV (4)
		Panel (a): Math		
Multigrade	0.16***	0.19***	0.20***	0.20***
	(0.04)	(0.06)	(0.05)	(0.06)
Class size	-0.01***	-0.01	-0.01***	-0.01**
	(0.00)	(0.01)	(0.00)	(0.01)
		Panel (b): Language		
Multigrade	0.11***	0.14**	0.15***	0.14***
	(0.04)	(0.06)	(0.05)	(0.05)
Class size	-0.01***	-0.01	-0.01***	-0.01***
	(0.00)	(0.01)	(0.00)	(0.00)
Instrumented variable(s)	Multigrade	Multigrade, Class size	Multigrade	Multigrade, Class size
Sample	Full	Full	Reduced	Reduced
Observations	92,504	92,504	32,659	32,659

IV estimates of the effect of multigrading on a child's test score. Dependent variable: Math test score (Panel a), Language test score (Panel b). All models include controls for child's gender, age, nationality, father's and mother's educational level, and father's and mother's profession. All models also include variables for altitude and population of the municipality, geographical macro-area, and road distance in time to the closest alternative school. Standard errors are clustered at the school level and reported in brackets. *, **, *** indicate statistical significance at the 10%, 5%, and 1% levels, respectively.

Table 7: Sensitivity Analysis

	IV (1)	IV (2)	IV (3)	IV (4)
	Panel (a): Inclusion of school size			
Multigrade	0.15***	0.18***	0.19***	0.19***
	(0.04)	(0.06)	(0.06)	(0.06)
Class size	-0.01***	-0.01	-0.01***	-0.01**
	(0.00)	(0.01)	(0.00)	(0.01)
	Panel (b): Regional (NUTS 2) FE			
Multigrade	0.17***	0.19***	0.20***	0.19***
	(0.04)	(0.06)	(0.05)	(0.06)
Class size	-0.01***	-0.01	-0.01***	-0.01***
	(0.00)	(0.01)	(0.00)	(0.00)
	Panel (c): Provincial (NUTS 3) FE			
Multigrade	0.19***	0.20***	0.22***	0.20***
	(0.04)	(0.06)	(0.05)	(0.06)
Class size	-0.01***	-0.01	-0.01***	-0.02***
	(0.00)	(0.01)	(0.00)	(0.00)
	Panel (d): Parents' missing information			
Multigrade	0.15***	0.15**	0.21***	0.20***
	(0.04)	(0.06)	(0.05)	(0.06)
Class size	-0.01***	-0.01	-0.01***	-0.02***
	(0.00)	(0.01)	(0.00)	(0.01)
Instrumented variable(s)	Multigrade	Multigrade, Class size	Multigrade	Multigrade, Class size
Sample	Full	Full	Reduced	Reduced

Sensitivity analysis for baseline estimates. Dependent variable: Combined Math-Language test score. All models include controls for child's gender, age, nationality, father's and mother's educational level, and father's and mother's profession. All models also include variables for altitude and population of the municipality, geographical macro-area (except Panels b and c), and road distance in time to the closest alternative school. Standard errors are clustered at the school level and reported in brackets. *, **, *** indicate statistical significance at the 10%, 5%, and 1% levels, respectively.

Table 8: Multigrading and Child Achievement: The Case of Northern Regions

	Combined Math-Language					
	OLS (1)	IV (2)	IV (3)	OLS (4)	IV (5)	IV (6)
Multigrade	0.08**	0.10**	0.15**	0.08**	0.13**	0.11*
	(0.04)	(0.05)	(0.07)	(0.04)	(0.05)	(0.06)
Class size	-0.01***	-0.01***	-0.00	-0.01**	-0.01**	-0.01*
	(0.00)	(0.00)	(0.01)	(0.00)	(0.00)	(0.01)
Instrumented variable(s)		Multigrade	Multigrade, Class size		Multigrade	Multigrade, Class size
Sample	Full Only North	Full Only North	Full Only North	Reduced Only North	Reduced Only North	Reduced Only North
Observations	58,345	58,345	58,345	18,277	18,277	18,277

IV estimates of the effect of multigrading on a child's test score. Dependent variable: Combined Math-Language test score. The analysis is only based on northern Italian regions. All models include controls for child's gender, age, nationality, father's and mother's educational level, and father's and mother's profession. All models also include variables for altitude and population of the municipality, and road distance in time to the closest alternative school. Standard errors are clustered at the school level and reported in brackets. *, **, *** indicate statistical significance at the 10%, 5%, and 1% levels, respectively.

Table 9: Heterogeneous Effects of Multigrading on Child Achievement

	Child's gender		Parental education	
	IV	IV	IV	IV
	(1)	(2)	(3)	(4)
	Female		No one with university	
Multigrade	0.18***	0.23***	0.16***	0.21***
	(0.05)	(0.05)	(0.04)	(0.05)
Class size	-0.01***	-0.01***	-0.01***	-0.01***
	(0.00)	(0.00)	(0.00)	(0.00)
Instrumented variable(s)	Multigrade	Multigrade	Multigrade	Multigrade
Sample	Full	Full	Reduced	Reduced
Observations	45,243	15,910	79,889	28,601
	Male		One with university	
Multigrade	0.12***	0.16***	0.10	0.11
	(0.05)	(0.05)	(0.06)	(0.07)
Class size	-0.01***	-0.01***	-0.01***	-0.01***
	(0.00)	(0.00)	(0.00)	(0.00)
Instrumented variable(s)	Multigrade	Multigrade	Multigrade	Multigrade
Sample	Full	Reduced	Full	Reduced
Observations	47,261	16,749	12,615	4,058

Heterogeneous analysis by child's gender (columns 1 and 2) and parental background (columns 3 and 4). Dependent variable: Combined Math-Language test score. All models include controls for child's gender, age, nationality, father's and mother's educational level, and father's and mother's profession. All models also include variables for altitude and population of the municipality, geographical macro-area, and road distance in time to the closest alternative school. Standard errors are clustered at the school level and reported in brackets. *, **, *** indicate statistical significance at the 10%, 5%, and 1% levels, respectively.

Table 10: Grade Composition and Child Achievement

	Combined Math-Language					
	Older Peers OLS (1)	Younger Peers OLS (2)	Mixed Peers OLS (3)	Older Peers OLS (4)	Younger Peers OLS (5)	Mixed Peers OLS (6)
Multigrade	0.16***	0.04	0.19*	0.16***	0.05	0.23**
	(0.05)	(0.04)	(0.10)	(0.06)	(0.05)	(0.10)
Class size	-0.01***	-0.01***	-0.01***	-0.01***	-0.01***	-0.01***
	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)
Sample	Full	Full	Full	Reduced	Reduced	Reduced
Observations	88,463	89,496	87,680	28,618	29,651	27,835

Analysis of the effect of multigrading according to class composition in terms of grades. Dependent variable: Combined Math-Language test score. Older peers means that children of higher grades (third, fourth and fifth grades) attend the same multigrade class of second-grade students. Younger peers means that only first graders attend the same multigrade class of second-grade students. Mixed peers means that both children of higher grades and first graders attend the same multigrade class of second-grade students. All models include controls for child's gender, age, nationality, father's and mother's educational level, and father's and mother's profession. All models also include variables for altitude and population of the municipality, geographical macro-area, and road distance in time to the closest alternative school. Standard errors are clustered at the school level and reported in brackets. *, **, *** indicate statistical significance at the 10%, 5%, and 1% levels, respectively.

Table 11: Multigrading and Child Achievement: the Case of Fifth-Grade Students

	Combined Math-Language					
	OLS (1)	IV (2)	IV (3)	OLS (4)	IV (5)	IV (6)
Multigrade	-0.02	0.01	-0.02	-0.02	-0.01	0.02
	(0.03)	(0.04)	(0.05)	(0.03)	(0.04)	(0.05)
Class size	-0.01**	-0.01**	-0.01*	-0.01***	-0.01***	-0.01*
	(0.00)	(0.00)	(0.01)	(0.00)	(0.00)	(0.00)
Instrumented variable(s)		Multigrade	Multigrade, Class size		Multigrade	Multigrade, Class size
Sample	Full	Full	Full	Reduced	Reduced	Reduced
Observations	89,780	89,780	89,780	31,155	31,155	31,155

OLS and IV estimates of the effect of multigrading on a child's (fifth-grade student) test score. Dependent variable: Combined Math-Language test score. All models include controls for child's gender, age, nationality, father's and mother's educational level, and father's and mother's profession. All models also include variables for altitude and population of the municipality, geographical macro-area, and road distance in time to the closest alternative school. Standard errors are clustered at the school level and reported in brackets. *, **, *** indicate statistical significance at the 10%, 5%, and 1% levels, respectively.

Appendix: Additional Material

A.1 Data Construction Process

In this Appendix we describe the process we used to: a) identify students attending multigrade classes in the INVALSI data, and b) identify the grade composition of multigrade classes.

a) Students in multigrade classes

As mentioned in the paper, no information about the class attended by students is available in the INVALSI data, so students attending a multigrade cannot be directly identified. To obtain this information, we merged three administrative archives. The first data set (the INVALSI data from now on) contains information about children's performance on the INVALSI test in school year 2012/2013. For each student, the test score in both mathematics and language as well as background information such as gender, age, nationality, attendance of preparatory schools, and parents' education and profession are available. Neither school names nor school characteristics and location are available in this data set. However, each individual record also includes a class and a school code, as well as geographical and demographic information about the municipality where the student's school is located. This piece of information is fundamental for our matching procedure and includes: i) the province where the school is located, ii) the population (in the 2001 and the 2011 census) of the municipality, iii) the size (in square km) of the municipality, and iv) the altitude of the municipality where the school is located.

A second administrative data set (School Register data from now on) provided by the Italian Ministry of Education contains detailed information about the characteristics of each Italian primary school in school year 2012/2013. All the Italian regions are covered in this data with the exception of Valle d'Aosta and Trentino Alto Adige. The School Register includes information such as school name, municipality, number of students (total and in each grade), number of classes (total and in each grade), and number of multigrade classes. Based on this information, we analyzed all of the possible combinations of grade composition at the school level to identify different types of schools. For example, if a school shows a positive number of second-grade students, but no second year single-grade classes, and at least one multigrade class, we can assume that second-grade students attend a multigrade class. We ended up with: i) schools where second-grade students attend a multigrade class; ii) schools where second-grade students attend one second year single-grade class; iii) schools where second-grade students attend more than one second year single-grade class; and iv) schools with no second-grade students. Note that we found no evidence of primary schools with both single and multigrade classes for the same grade.

Unfortunately, the INVALSI data and the School Register data cannot be matched directly. In fact, the first data set only identifies each primary school with an anonymous code. The only way to overcome this problem is to identify (at least) the names of the municipalities where the schools included in the INVALSI data set are located. Once identified, it would be possible to match the data set with the School Register, with municipality as the matching variable.

The Municipality Register data set provided by ISTAT is the last piece of information needed to complete the data construction process. The Municipality Register contains geographical and demographic information for each Italian municipality. This information (province, population in the 2001 and 2011 census, size and altitude of the municipality) is the same as that contained in the INVALSI data, therefore making the merger of the INVALSI data set with the Municipality register data set possible. We use geographical and demographic information as key identifying variables in the matching process to obtain the INVALSI+ISTAT data.

The last step is the matching of the INVALSI+ISTAT data with the School Register data based on municipality names. Unfortunately, with this last matching, we are able to uniquely identify only schools located in municipalities hosting no more than one school. We repeated the same procedure to obtain the data for fifth-grade students.

b) Grade composition of multigrade classes

As mentioned, in the paper, no data identify the grade composition of multigrade classes. To get to this result, we use the data built in the previous paragraph and apply a wide set of rules to identify grade composition of multigrade classes. These rules are based on the information originally included in the School Register. For example, we define the following *Rule 1* to identify a multigrade class whose students are first and second graders only (therefore second graders are the older peers in the multigrade class). According to the *Rule 1* the school has:

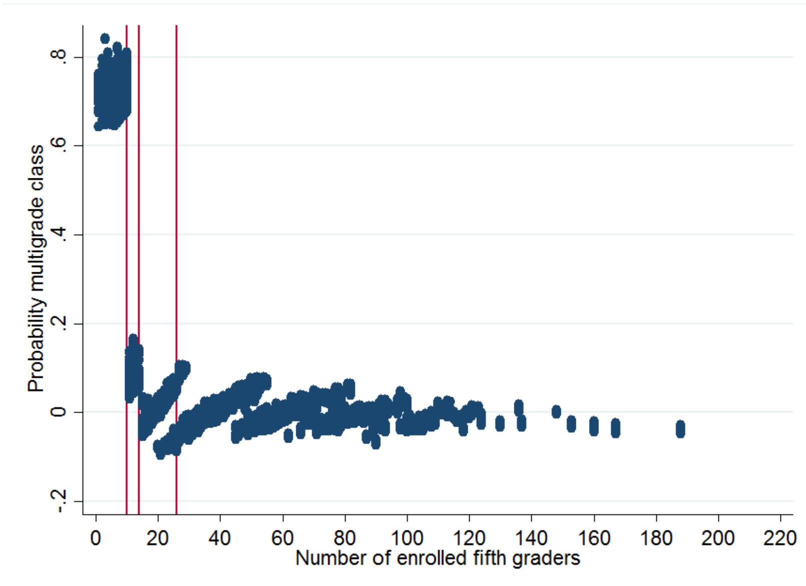
- a) one multigrade class;
- b) no first- and second-grade single classes;
- c) first- and second-grade students;
- d) third-, fourth- and fifth-grade single classes;
- e) third-, fourth- and fifth-grade students.

We elaborate about 40 such rules to enumerate all the possible combinations of students of different grades and to describe the classes in our data.

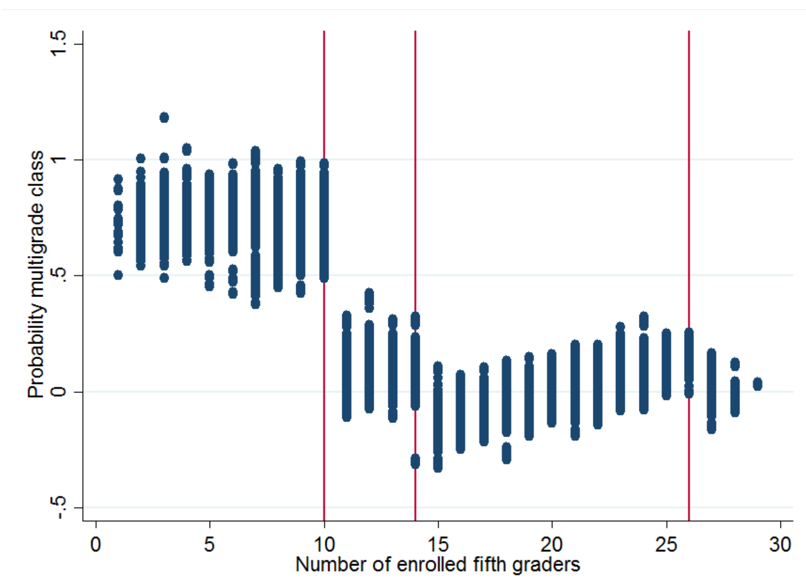
A.2 Additional Figures and Tables

Figure A.1: Number of Fifth Graders and Individual Probability of Multigrading

(a) *Full sample*



(b) *Reduced sample*



Notes: This figure shows the predicted individual probability of ending up in a multigrade class for fifth-grade students as a function of the number of fifth-grade students enrolled in a school. Panel (a) refers to the full sample, Panel (b) is based on the reduced sample. The predicted individual probability of attending a multigrade class (y-axis) is obtained through first-stage estimates in Table A.2, columns (1) and (4). Refer to the text and to Table A.2 for further details about the empirical model underlying this figure.

Table A.1: Summary Statistics (Fifth Grade)

	Full sample		Reduced sample	
	Mean (1)	St.Dev. (2)	Mean (3)	St.Dev. (4)
Math	27.70	8.94	27.94	8.97
Language	62.82	12.48	63.04	12.52
Multigrade	0.05	0.22	0.15	0.35
Class size	18.86	4.00	17.08	4.71
Age	9.98	0.32	9.97	0.33
Female	0.50	0.50	0.50	0.50
Italian	0.89	0.31	0.91	0.29
Migrant 1st gen.	0.04	0.21	0.04	0.21
Migrant 2nd gen.	0.06	0.24	0.05	0.22
Northwest	0.44	0.50	0.41	0.49
Northeast	0.17	0.37	0.24	0.42
Central area	0.12	0.32	0.12	0.33
South	0.20	0.40	0.24	0.42
Islands	0.08	0.27	0.09	0.29
Time distance (min.)	5.58	3.37	6.25	3.70
Population (2011)	4,609	3,057	1,919	823
Altitude	268	227	364	255
Father university	0.06	0.24	0.06	0.23
Father high school	0.76	0.43	0.77	0.42
Father other	0.18	0.38	0.17	0.38
Mother university	0.08	0.28	0.08	0.27
Mother high school	0.75	0.43	0.76	0.42
Mother other	0.17	0.37	0.16	0.37
Observations	89,780		31,155	

This table shows summary statistics for the samples analyzed in this work. The table reports data about fifth-grade students. Columns (1) and (2) refer to the full sample, while columns (3) and (4) refer to the reduced sample.

Table A.2: First-Stage Estimates (Fifth Grade)

	Full sample			Reduced sample		
	Model (1)	Model (2)		Model (1)	Model (2)	
	Multigrade	Multigrade	Class size	Multigrade	Multigrade	Class size
	(1)	(2)	(3)	(4)	(5)	(6)
$5^{th} Graders \leq 10$	0.76***	0.71***	-5.19***	1.07***	0.65***	-12.86***
	(0.01)	(0.02)	(0.20)	(0.02)	(0.02)	(0.31)
$11 \leq 5^{th} Graders \leq 14$	0.11***	0.06***	-4.85***	0.44***	0.02*	-12.89***
	(0.01)	(0.01)	(0.16)	(0.03)	(0.01)	(0.27)
$15 \leq 5^{th} Graders \leq 26$	-0.01***	-0.00*	1.09***	0.22***	-0.01	-6.95***
	(0.00)	(0.00)	(0.18)	(0.02)	(0.01)	(0.25)
Class size	0.01***			0.03***		
	(0.00)			(0.00)		
Father high school	0.00	0.00	0.01	0.00	0.01	0.08
	(0.00)	(0.00)	(0.05)	(0.01)	(0.01)	(0.08)
Mother high school	0.00**	0.00**	0.06	0.01**	0.01**	0.04
	(0.00)	(0.00)	(0.04)	(0.01)	(0.01)	(0.07)
Instrumented variable(s)	Multigrade	Multigrade+Class size		Multigrade	multigrade+Class size	
SW Chi-sq. (UId)	> 100	> 100	> 100	> 100	> 100	> 100
P-value	0.00	0.00	0.00	0.00	0.00	0.00
SW F (WId)	> 100	> 100	> 100	> 100	> 100	> 100
P-value	0.00	0.00	0.00	0.00	0.00	0.00
KP (WId)	> 100	> 100	> 100	> 100	> 100	> 100
Observations	89,780	89,780	89,780	31,155	31,155	31,155

First-stage estimates. Dependent variable: Being enrolled in a multigrade class (columns 1,2,4, and 5), class size (columns 3 and 6). The reference category for the number of second graders is the class $5^{th} Graders > 26$. The reference category for father's and mother's education is completed university. All models include controls for child's gender, age, nationality, and father's and mother's profession. All models also include variables for altitude and population of the municipality, geographical macro-area, and road distance in time to the closest alternative school. Standard errors are clustered at the school level and reported in brackets. *, **, *** indicate statistical significance at the 10%, 5%, and 1% levels, respectively.

Table A.3: Distance to the Closest School as Instrument for Parents' Preferences (Fifth Grade)

	Combined Math-Language			
	IV (1)	IV (2)	IV (3)	IV (4)
Multigrade	0.00	-0.03	-0.01	0.01
	(0.04)	(0.05)	(0.04)	(0.05)
Class size	-0.01**	-0.01*	-0.01***	-0.01
	(0.00)	(0.01)	(0.00)	(0.00)
Instrumented variable(s)	Multigrade	Multigrade, Class size	Multigrade	Multigrade, Class size
Sample	Full	Full	Reduced	Reduced
Observations	89,780	89,780	31,155	31,155

IV estimates of the effect of multigrading on a child's (fifth-grade student) test score. Dependent variable: Combined Math-Language test score. Road distance in time to the closest alternative school is used as additional instrument for being enrolled in a multigrade class and class size. All models include controls for child's gender, age, nationality, father's and mother's educational level, and father's and mother's profession. All models also include variables for altitude and population of the municipality, geographical macro-area. Standard errors are clustered at the school level and reported in brackets. *, **, *** indicate statistical significance at the 10%, 5%, and 1% levels, respectively.

Table A.4: Single Test Scores (Fifth Grade)

	IV (1)	IV (2)	IV (3)	IV (4)
	Panel (a): Math			
Multigrade	0.02	-0.01	-0.00	0.03
	(0.04)	(0.06)	(0.05)	(0.06)
Class size	-0.01**	-0.01*	-0.01***	-0.01*
	(0.00)	(0.01)	(0.00)	(0.01)
	Panel (b): Language			
Multigrade	-0.00	-0.02	-0.02	-0.00
	(0.03)	(0.05)	(0.04)	(0.05)
Class size	-0.00**	-0.01	-0.01***	-0.01
	(0.00)	(0.00)	(0.00)	(0.00)
Instrumented variable(s)	Multigrade	Multigrade, Class size	Multigrade	Multigrade, Class size
Sample	Full	Full	Reduced	Reduced
Observations	89,780	89,780	31,155	31,155

IV estimates of the effect of multigrading on a child's (fifth-grade student) test score. Dependent variable: Math test score (Panel a), Language test score (Panel b). All models include controls for child's gender, age, nationality, father's and mother's educational level, and father's and mother's profession. All models also include variables for altitude and population of the municipality, geographical macro-area, and road distance in time to the closest alternative school. Standard errors are clustered at the school level and reported in brackets. *, **, *** indicate statistical significance at the 10%, 5%, and 1% levels, respectively.

Table A.5: Multigrading and Child Achievement: The Case of Northern Regions (Fifth Grade)

	Combined Math-Language					
	OLS (1)	IV (2)	IV (3)	OLS (4)	IV (5)	IV (6)
Multigrade	-0.05	-0.03	-0.01	-0.06	-0.06	-0.00
	(0.03)	(0.04)	(0.06)	(0.04)	(0.05)	(0.06)
Class size	-0.01**	-0.01**	-0.00	-0.01**	-0.01**	0.00
	(0.00)	(0.00)	(0.01)	(0.00)	(0.00)	(0.01)
Instrumented variable(s)		Multigrade	Multigrade, Class size		Multigrade	Multigrade, Class size
Sample	Full Only North	Full Only North	Full Only North	Reduced Only North	Reduced Only North	Reduced Only North
Observations	54,402	54,402	54,402	17,131	17,131	17,131

IV estimates of the effect of multigrading on a child's (Fifth-grade student) test score. Dependent variable: Combined Math-Language test score. The analysis is only based on northern Italian regions. All models include controls for child's gender, age, nationality, father's and mother's educational level, and father's and mother's profession. All models also include variables for altitude and population of the municipality, and road distance in time to the closest alternative school. Standard errors are clustered at the school level and reported in brackets. *, **, *** indicate statistical significance at the 10%, 5%, and 1% levels, respectively.

Working Paper del Dipartimento di Economia e Finanza

1. L. Colombo, H. Dawid, *Strategic Location Choice under Dynamic Oligopolistic Competition and Spillovers*, novembre 2013.
2. M. Bordignon, M. Gamalerio, G. Turati, *Decentralization, Vertical Fiscal Imbalance, and Political Selection*, novembre 2013.
3. M. Guerini, *Is the Friedman Rule Stabilizing? Some Unpleasant Results in a Heterogeneous Expectations Framework*, novembre 2013.
4. E. Brenna, C. Di Novi, *Is caring for elderly parents detrimental to women's mental health? The influence of the European North-South gradient*, novembre 2013.
5. F. Sobbrio, *Citizen-Editors' Endogenous Information Acquisition and News Accuracy*, novembre 2013.
6. P. Bingley, L. Cappellari, *Correlation of Brothers Earnings and Intergenerational Transmission*, novembre 2013.
7. T. Assenza, W. A. Brock, C. H. Hommes, *Animal Spirits, Heterogeneous Expectations and the Emergence of Booms and Busts*, dicembre 2013.
8. D. Parisi, *Is There Room for 'Fear' as a Human Passion in the Work by Adam Smith?*, gennaio 2014.
9. E. Brenna, F. Spandonaro, *Does federalism induce patients' mobility across regions? Evidence from the Italian experience*, febbraio 2014.
10. A. Monticini, F. Ravazzolo, *Forecasting the intraday market price of money*, febbraio 2014.
11. Tiziana Assenza, Jakob Grazzini, Cars Hommes, Domenico Massaro, *PQ Strategies in Monopolistic Competition: Some Insights from the Lab*, marzo 2014.
12. R. Davidson, A. Monticini, *Heteroskedasticity-and-Autocorrelation-Consistent Bootstrapping*, marzo 2014.
13. C. Lucifora, S. Moriconi, *Policy Myopia and Labour Market Institutions*, giugno 2014.
14. N. Pecora, A. Spelta, *Shareholding Network in the Euro Area Banking Market*, giugno 2014.
15. G. Mazzolini, *The economic consequences of accidents at work*, giugno 2014.
16. M. Ambrosanio, P. Balduzzi, M. Bordignon, *Economic crisis and fiscal federalism in Italy*, settembre 2014.
17. P. Bingley, L. Cappellari, K. Tatsiramos, *Family, Community and Long-Term Earnings Inequality*, ottobre 2014.
18. S. Frazzoni, M. L. Mancusi, Z. Rotondi, M. Sobrero, A. Vezzulli, *Innovation and export in SMEs: the role of relationship banking*, novembre 2014.
19. H. Gnutzmann, *Price Discrimination in Asymmetric Industries: Implications for Competition and Welfare*, novembre 2014.
20. A. Baglioni, A. Boitani, M. Bordignon, *Labor mobility and fiscal policy in a currency union*, novembre 2014.
21. C. Nielsen, *Rational Overconfidence and Social Security*, dicembre 2014.
22. M. Kurz, M. Motolese, G. Piccillo, H. Wu, *Monetary Policy with Diverse Private Expectations*, febbraio 2015.
23. S. Piccolo, P. Tedeschi, G. Ursino, *How Limiting Deceptive Practices Harms Consumers*, maggio 2015.
24. A.K.S. Chand, S. Currarini, G. Ursino, *Cheap Talk with Correlated Signals*, maggio 2015.
25. S. Piccolo, P. Tedeschi, G. Ursino, *Deceptive Advertising with Rational Buyers*, giugno 2015.

26. S. Piccolo, E. Tarantino, G. Ursino, *The Value of Transparency in Multidivisional Firms*, giugno 2015.
27. G. Ursino, *Supply Chain Control: a Theory of Vertical Integration*, giugno 2015.
28. I. Aldasoro, D. Delli Gatti, E. Faia, *Bank Networks: Contagion, Systemic Risk and Prudential Policy*, luglio 2015.
29. S. Moriconi, G. Peri, *Country-Specific Preferences and Employment Rates in Europe*, settembre 2015.
30. R. Crinò, L. Ogliari, *Financial Frictions, Product Quality, and International Trade*, settembre 2015.
31. J. Grazzini, A. Spelta, *An empirical analysis of the global input-output network and its evolution*, ottobre 2015.
32. L. Cappellari, A. Di Paolo, *Bilingual Schooling and Earnings: Evidence from a Language-in-Education Reform*, novembre 2015.
33. A. Litina, S. Moriconi, S. Zanaj, *The Cultural Transmission of Environmental Preferences: Evidence from International Migration*, novembre 2015.
34. S. Moriconi, P. M. Picard, S. Zanaj, *Commodity Taxation and Regulatory Competition*, novembre 2015.
35. M. Bordignon, V. Grembi, S. Piazza, *Who do you blame in local finance? An analysis of municipal financing in Italy*, dicembre 2015.
36. A. Spelta, *A unified view of systemic risk: detecting SIFIs and forecasting the financial cycle via EWSs*, gennaio 2016.
37. N. Pecora, A. Spelta, *Discovering SIFIs in interbank communities*, febbraio 2016.
38. M. Botta, L. Colombo, *Macroeconomic and Institutional Determinants of Capital Structure Decisions*, aprile 2016.
39. A. Gamba, G. Immordino, S. Piccolo, *Organized Crime and the Bright Side of Subversion of Law*, maggio 2016.
40. L. Corno, N. Hildebrandt, A. Voena, *Weather Shocks, Age of Marriage and the Direction of Marriage Payments*, maggio 2016.
41. A. Spelta, *Stock prices prediction via tensor decomposition and links forecast*, maggio 2016.
42. T. Assenza, D. Delli Gatti, J. Grazzini, G. Ricchiuti, *Heterogeneous Firms and International Trade: The role of productivity and financial fragility*, giugno 2016.
43. S. Moriconi, *Taxation, industry integration and production efficiency*, giugno 2016.
44. L. Fiorito, C. Orsi, *Survival Value and a Robust, Practical, Joyless Individualism: Thomas Nixon Carver, Social Justice, and Eugenics*, luglio 2016.
45. E. Cottini, P. Ghinetti, *Employment insecurity and employees' health in Denmark*, settembre 2016.
46. G. Cecere, N. Corrocher, M. L. Mancusi, *Financial constraints and public funding for eco-innovation: Empirical evidence on European SMEs*, settembre 2016.
47. E. Brenna, L. Gitto, *Financing elderly care in Italy and Europe. Is there a common vision?*, settembre 2016.
48. D. G. C. Britto, *Unemployment Insurance and the Duration of Employment: Theory and Evidence from a Regression Kink Design*, settembre 2016.
49. E. Caroli, C.Lucifora, D. Vigani, *Is there a Retirement-Health Care utilization puzzle? Evidence from SHARE data in Europe*, ottobre 2016.
50. G. Femminis, *From simple growth to numerical simulations: A primer in dynamic programming*, ottobre 2016.
51. C. Lucifora, M. Tonello, *Monitoring and sanctioning cheating at school: What works? Evidence from a national evaluation program*, ottobre 2016.

52. A. Baglioni, M. Esposito, *Modigliani-Miller Doesn't Hold in a "Bailinable" World: A New Capital Structure to Reduce the Banks' Funding Cost*, novembre 2016.
53. L. Cappellari, P. Castelnovo, D. Checchi, M. Leonardi, *Skilled or educated? Educational reforms, human capital and earnings,* novembre 2016.
54. D. Britto, S. Fiorin, *Corruption and Legislature Size: Evidence from Brazil*, dicembre 2016.
55. F. Andreoli, E. Peluso, *So close yet so unequal: Reconsidering spatial inequality in U.S. cities*, febbraio 2017.
56. E. Cottini, P. Ghinetti, *Is it the way you live or the job you have? Health effects of lifestyles and working conditions,* marzo 2017.
57. A. Albanese, L. Cappellari, M. Leonardi, *The Effects of Youth Labor Market Reforms: Evidence from Italian Apprenticeships*; maggio 2017.
58. S. Perdichizzi, *Estimating Fiscal multipliers in the Eurozone. A Nonlinear Panel Data Approach*, maggio 2017.
59. S. Perdichizzi, *The impact of ECBs conventional and unconventional monetary policies on European banking indexes returns*, maggio 2017.
60. E. Brenna, *Healthcare tax credits: financial help to taxpayers or support to higher income and better educated patients? Evidence from Italy*, giugno 2017.
61. G. Gokmen, T. Nannicini, M. G. Onorato, C. Papageorgiou, *Policies in Hard Times: Assessing the Impact of Financial Crises on Structural Reforms*, settembre 2017.
62. M. Tettamanzi, *E Many Pluribus Unum: A Behavioural Macro-Economic Agent Based Model*, novembre 2017.
63. A. Boitani, C. Punzo, *Banks' leverage behaviour in a two-agent New Keynesian model*, gennaio 2018.
64. M. Bertoni, G. Brunello, L. Cappellari, *Parents, Siblings and Schoolmates. The Effects of Family-School Interactions on Educational Achievement and Long-term Labor Market Outcomes,* gennaio 2018.
65. G. P. Barbetta, G. Sorrenti, G. Turati, *Multigrading and Child Achievement,* gennaio 2018.