



Working Paper n. 03 – 2018

THE EFFECT OF MULTIGRADE CLASSES ON COGNITIVE AND NON- COGNITIVE SKILLS. CAUSAL EVIDENCE EXPLOITING MINIMUM CLASS SIZE RULES IN ITALY

Daniele Checchi

Dipartimento di Economia, Management e
Metodi Quantitativi
Università degli Studi di Milano
Via Festa del Perdono, 7
20122 Milano
IZA – Institute of Labor Economics, Bonn
e-mail: daniele.checchi@unimi.it

Maria De Paola

Dipartimento di Economia, Statistica e Finanza
“*Giovanni Anania*” - DESF
Università della Calabria
Ponte Pietro Bucci, Cubo 1/C
Tel.: +39 0984 492459
Fax: +39 0984 492421
e-mail: m.depaola@unical.it

Ottobre 2018



The Effect of Multigrade Classes on Cognitive and non-Cognitive Skills. Causal Evidence Exploiting Minimum Class Size Rules in Italy*

Daniele Checchi

(University of Milan and IZA)

Maria De Paola

(Department of Economics, Statistics and Finance, University of Calabria and IZA)

Abstract. We analyse how schooling in multigrade classes affects the formation of student cognitive and non-cognitive skills. Our identification strategy is based on some institutional features of the Italian educational system establishing a minimum number of students per class. Classes that do not reach the minimum number of pupils are organized in multigrade classes. In addition, the Italian law also establishes a maximum number of students for multigrade classes, which implies that class size in multigrade classes is very similar to class size in small single grade classes with a number of students just above the minimum size. Using census data on 5th grade Italian students, we find that pupils in multigrade classrooms obtain worse test scores both in literacy and numeracy standardized tests compared to comparable pupils in single grade classroom. While the effect is small and not always statistically significant for the literacy score, we find a large and highly statistically significant effect on the numeracy score. We also find that pupils placed in multigrade classes tend to have a more external centred locus of control. Our results are robust to different specifications including controls for class size and a number of student and school characteristics.

JEL classification: I21, I28, C36

Keywords: multigrade classes, mixed-age classes, cognitive skills, non-cognitive skills,

*The authors would like to thank two anonymous referees, Colin Green, Marco Paccagnella, Vincenzo Scoppa, Diane Whitmore Schanzenbach and seminar/conference participants at the 2016 Invalsi meeting “I dati INVALSI: uno strumento per la ricerca” (Rome, 20/9/2016), at the Ninth International Workshop on Applied Economics of Education (Catanzaro, June, 2018) and at Trento University (May, 2018) for useful comments and suggestions. We also thanks Gianna Barbieri (MIUR Statistical office) for providing institutional data, and Patrizia Falzetti (INVALSI) for kindly merging this information with the public use file of students’ test scores.

1. Introduction

Multigrade classes, where students from two (or more) adjacent grades are grouped within one classroom, are common in many developing and developed countries. According to UNESCO (2005 Agenda for Educational Planning) approximately one third of all classes across the world are multigrade classes. In 2007, about 28% of schools in the United States were adopting this type of educational practice when the number of pupils was too small. The incidence of multigrade classes is also high in many European countries, especially in less populated areas. For instance, in France about 37 per cent of primary school pupils are in such classes. In Finland and in the Netherlands multigrade classes prevail over single grade ones (Mulkeen and Higgings, 2009; Lindström and Lindahl, 2011).

The use of multigrade classes often responds to the need of providing school services at the student's place of residence at a reasonable cost. In fact, in many developed countries the presence of this type of school organization is typical of rural or mountainous areas that in recent years, especially in some countries, experienced a drastic reduction in resident population: in these circumstances multigrade classes allow schools to remain located closer to the families they serve since there are not enough children to fill a conventional single grade class.

However, cost-saving considerations are to be evaluated against the effects that multigrade classes may produce on student outcomes. From a theoretical point of view, the effects produced by grade-mixing on student achievement can be either positive or negative. For instance, on the one hand, such a diverse environment in terms of age, skills and maturity can foster cognitive skills; on the other, the fact that teachers are forced to jump from one program to another and to interact with pupils with different needs or skills might reduce the effectiveness of their teaching.

The empirical literature trying to figure out which of the effects produced by multigrade classes is more relevant and to understand if students of a mixed-grade classroom are actually advantaged or at a disadvantage is scarce. This is mainly due to non-random selection into multigrade classes. For instance, schools might be more likely to adopt multigrade classes if they expect to obtain better results or if the teaching staff is more sympathetic with this type of educational practice. Teachers might try to avoid multigrade classes because of the higher effort required by teaching different programs and selection might not be random. In addition, the assignment of students to multigrade classes might depend on their unobservable characteristics. The few studies that have tried to solve these problems with appropriate techniques led to mixed results.¹ Sims (2008) uses an instrumental variable strategy based on class size caps imposed by the California Class Size Reduction Program and shows that multigrade classes negatively affect test scores in Grades 2 and 3. Negative effects both on final grade attainment and labour market participation are also found by Gerhardt et al. (2016) who exploit a natural experiment deriving from the abolition of

¹A review of earlier studies is provided by Veenman (1995) who surveys 56 papers and concludes that pupils in multigrade classrooms show results that are similar to those reached by pupils in classrooms that track pupils by grade. However, as pointed out by Mason and Burns (1997), many of these studies do not address sorting of pupils and teachers into multigrade classes.

parochial schools in Germany. Instead, Thomas (2012), adopting a school fixed effects method, finds that first graders are not harmed by being in a multigrade classes. Finally, Leuven and Ronning (2014), exploiting discontinuous grade mixing rules in Norwegian junior high schools, show that the presence of younger peers decreases achievement, while the reverse occurs in case of older peers. Even more scant is the evidence on the effects of multigrade classes on non-cognitive skills. The only paper looking at non-cognitive skills is Sattari (2016) who shows that placing students in multigrade classrooms induces more behavioural problems.

Our analysis contributes to this literature by providing additional evidence of the effects of multigrade classes both on cognitive and non-cognitive skills. At this aim we use a very rich dataset covering the entire populations of students attending the 5th grade in Italian public schools and providing information on students' performance in standardized test scores, grades assigned by teachers and on a measure of locus of control. Our identification strategy takes advantage of some institutional features of the Italian school legislation. According to existing rules (DM 331/98) primary school classes must consist of no less than 10 children² and must not exceed a maximum of 25 pupils per class. Classes that do not respect the minimum number of pupils are organized in multigrade classes. However, according to the Italian law, the possibility to form a multigrade class depends not only on the number of students enrolled in a given grade, but is also related to the number of students enrolled in adjacent grades. In fact, the law establishes that multigrade classes are subject to a maximum number of 12 students per class (with a minimum of 6). This further constraint implies that class size in multigrade classes is very similar to class size in small single grade classes with a number of students just above the cut-off point of 10. Exploiting these rules we build a binary multigrade predictor taking the value of 1 when two conditions are met: 1) the number of students enrolled in the grade is smaller than 10; 2) the total number of students enrolled in adjacent grades is smaller than 13.³ Even if these rules are not strictly respected (since school principal had some margin for flexibility) they represent a source of exogenous variation that we exploit to identify the effect of multigrade classes.

To minimize problems related to endogeneity in class size, we only focus on small schools that have no more than a classroom per grade. In these schools class size can be considered exogenous and mainly determined by variation in cohort size. Since there is only one classroom per grade there is little room for parents' and school administrators' choices and variation in class size can be considered as mainly related to natural randomness in population (Hoxby, 2000). Moreover, to compare classes of similar size we restrict our analysis to classes with no more than 13 students and no less than 5 students. This restriction, as we show in the Appendix of the paper, does not affect our results that remain substantially unchanged also when considering the whole sample, but reassures us that the estimated effect is not driven by differences in class size.

²A reform introduced in the school year 2009-10 increased the minimum size to 15 in hilly and plain areas (leaving the threshold of no less than 10 children in mountain areas and small islands). This reform was rolled out one grade per year, starting with grade 1. Students considered in our data remain subject to the old rules.

³ The second condition suggests that in each adjacent class there are 6 or 7 students per grade. As a consequence, a multigrade class can be formed by joining two adjacent classes while respecting the minimal and maximal thresholds.

We find a negative effect on students' performance both on literacy and numeracy standardized test scores. While the effect is not always statistically significant for the literacy score, a large and highly statistically significant effect is found on the numeracy score. Students placed in multigrade classes obtain a lower numeracy score than students in single grade classes by about half standard deviation. This effect holds true when we control for class size and for a number of students and school characteristics. It is worthwhile to notice that as the multigrade classes in our setting (see Section 2) are smaller in size than single grade classes, we would expect any possible negative effect produced by grade mixing to be counterbalanced by the positive effects deriving from working with a reduced number of students.

On the contrary, we find a negative but not always statistically significant effect when we look at grades assigned by teachers. This might depend on the fact that teachers' evaluations, in spite of standardized test scores, are the results of a more complex assessment process, which reflects the objective level of skills achieved by students, but also a number of other factors such as the perceived student effort, motivation, behaviour as well as parents' expectations (OECD, 2012; 2013). Then, teachers of multigrade classes, aware of the more complex environment faced by their students, possibly reward more generously their effort. This is found to be especially true for teachers working in the Southern part of the country, while for schools located in the North multigrade classes reduce students' performance also when measured using teachers' assessment.

Thanks to a complementary questionnaire proposed to students taking the INVALSI test, we are also able to analyse the effect of multigrade classes on locus of control, a psychological trait that has received considerable attention both in the psychological and economic literature.⁴ Locus of control captures '*a generalised attitude, belief or expectancy regarding the nature of the causal relationship between one's own behaviour and its consequences*' (Rotter, 1966). Individuals who believe that life's outcomes are due to their own efforts have an internal locus of control, while those believing that outcomes are due to external factors (e.g. luck) have an external locus of control (Gatz and Karel, 1993). Locus of control has been shown to explain a wide range of social and economic outcomes, such as educational attainment, earnings, unemployment and job search behaviour, life satisfaction and health investments (see for instance, Caliendo et al., 2015; Cebi, 2007; Chiteji, 2010; Coleman and Deleire, 2003; Groves, 2005). In addition, individuals with an internal locus of control are more able to cope with unanticipated life events such as health shocks and unemployment (Schurer, 2011; Caliendo *et al.*, 2015). In line with an emerging literature that describes non-cognitive skills as resulting from educational attainment, parental investments and policy interventions (Almlund et al. 2011), we consider the impact of educational inputs on individual locus of control. We find that students placed in multigrade classes are more likely to have an external locus of control. However, the

⁴ Multigrade classes might also affect other non-cognitive abilities. Unfortunately the data at hand do not provide other measures, such as self-confidence or Big Five personality traits. In addition to the information allowing us to build the indicator of locus of control, the questionnaire includes some questions on students' self-assessment of their ability in numeracy and literacy. We have used this information to build different indicators of self-assessed ability and used as outcome variables and we do not find any statistical significant effect of multigrade classes.

effect is not robust and loses significance when, instead of focusing on small classes, we consider the whole sample of students.

Since our results pertain to students attending the 5th grade who are likely to have been in a multigrade class also in previous years of their educational process (unfortunately we do not have information on this), it is likely that the negative effect we find represents the cumulative effect of having attended a substantial fraction of primary school in a multigrade environment. This effect might hide differentiated effects according to peers' age. In fact, even if in our setting students attending the 5th grade in a multigrade classroom interact with younger mates, in previous years they are likely to have interacted also with older peers. As multigrade classes are typically formed combining adjacent grades, the peer group changes over the school cycle. More precisely, in Italy, it is quite common to have the first three grades (1st, 2nd and 3rd) grouped in a classroom and the last two (4th and 5th grade) in another classroom; then, a student who starts primary school in a multigrade class during her/his first year interacts with older peers but in the following years she/he finds himself in a class with younger peers. This implies that the effect we find is the sum of the effect of sharing the classroom with higher grade mates and the effect of sharing the classroom with lower grade ones.

The rest of the paper is organized as follows. Section 2 provides a description of the institutional setting of Italian schools and explains the rules followed to decide whether teaching activity will take place in a multigrade or in a single grade class. Section 3 presents our estimation strategy and discusses possible threats to the validity of our research design providing a number of checks on our first stage. Estimates of the effects of multigrade classes on cognitive skills are reported in Section 4, while in Section 5 we analyse the effects on locus of control. Section 6 presents some robustness checks. Section 7 concludes.

2. Institutional background

In 2015 about 51 thousand children (1.5% of the whole population of students enrolled in primary school) attended primary school in a multigrade classroom in Italy. The majority of these students lives in small municipalities that are at a considerable distance from main towns and offer only limited access to essential services such as education, mobility, health, etc. In recent years, these municipalities have suffered a strong population decline and demographic ageing which has led to smaller and smaller schools, with a low number of students per class (in 2016, in 13% of Italian municipalities the average number of pupils per class in primary school was less than 10). In these areas multigrade classes can be a cost-effective tool to retain locally provided education and to avoid to pupils and their families the cost associated to attending schools located in a different municipality.

Teachers in Italian primary schools are required to have obtained a university degree in Education and they have to teach an identical nation-wide curriculum, defined for each grade (they are supposed to follow the 5th grade curriculum with 5th graders, the 4th grade curriculum with the 4th graders, etc.). This implies that students are taught the same curriculum by teachers with similar qualification, irrespective of

whether they attend a single grade class or a multigrade one. Nonetheless, teachers might adjust their teaching techniques to the different environment and this could have an impact on the students' learning. Rules about the number of teachers in each class of primary education have changed over time; until 1990 there was a single teacher for each class, in the period going from 1990 to 2008 there were two teachers per class, while in the school year 2008/09 a new reform (*riforma Gelmini*) reintroduced the possibility to have a single teacher for each class (*maestro unico*). The cohort of students we consider in this study (enrolled at the first grade in 2007/08) experienced two different teachers specialized in the main subjects. Teaching organization is the same for multigrade and single grade classes.

The allocation of students to multigrade classes is decided following a mix of rules and discretion. In the Italian educational system, until 2008/09 school year, primary school classes were subject to a minimum size of 10 and a maximum of 25 (Decreto Ministeriale 331/98). In 2009-10 the minimum and the maximum were increased to 15 and 27 respectively (Decreto del Presidente della Repubblica 81/2009), with the exception of schools in mountain areas and small islands where the minimum number of pupils was maintained at 10. If the number of students in a specific grade was lower than that threshold, school heads were required to form multigrade classes (*pluriclassi*), grouping students of adjacent grades. The Italian law also established that multigrade classes cannot be composed by more than 12 students (changed to 18 students in 2009). The 2009 reform was rolled out one grade per year starting with grade 1 and students considered in our data were subject to the old rule. On these basis, we build a dummy variable *Predicted Multigrade* taking the value of one when the number of students enrolled in 5th grade is smaller than 10 and if the total number of students enrolled in grades 4 and 5 is smaller than 13. Since both the cap on the minimum number of students per class and the cap on the maximum number of students in multigrade classes could be applied with a discretionary margin of 10% above/below the numbers set by law, *Predicted Multigrade* does not perfectly predict student placement in a multigrade class.

The situation is depicted in Figure 1, while considering students in two adjacent grades, say 4th and 5th grade (in Italy primary school extends from grade 1 to grade 5). If the number of students in each age cohort exceeds 10, a standard single class is formed. When both cohorts fall below 10, there is room for forming a multigrade class among them, conditional on their sum being greater than 6 (otherwise they are to be mixed with students from previous grades – the small triangle at the origin). But their sum must not also exceed 12, because in such a case the multigrade class would be too large to be taught in an effective way. A case where there are 7 (or 8 or 9) students in each age cohort is ambiguous and we ignore which is the solution adopted by the school principal. For this reason, our measure of *Predicted multigrade* does not perfectly identify the type of teaching received by a specific student.

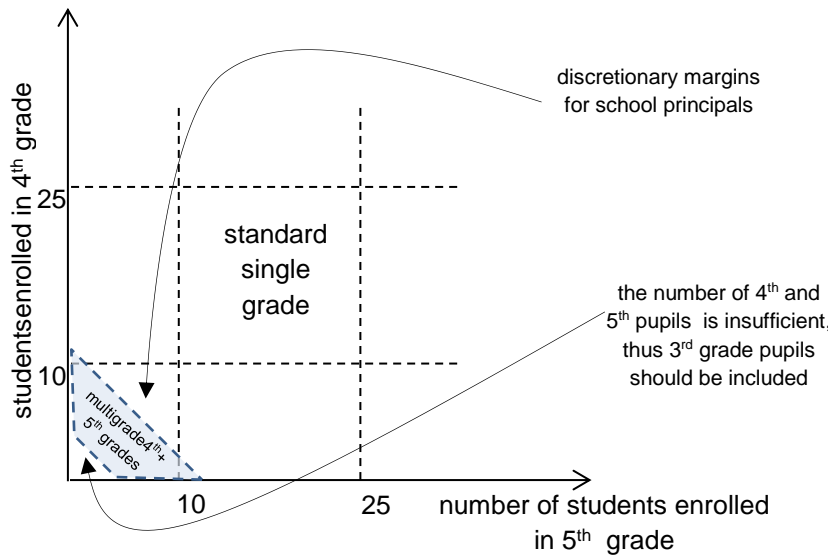


Figure 1. Multigrade class formation

3. Data

To investigate the effect of multigrade classes on student outcomes we rely on data from the Italian National Assessment Program, INVALSI, a government agency that carries out a yearly testing of student attainment in literacy and numeracy. The evaluation covers the entire population of students attending 2nd and 5th grade (primary school), as well as 8th and 10th graders (lower and upper secondary schools respectively). The dataset provides information not only on standardized test score results but also on the marks assigned by math and language teachers.

In our work we focus on primary schools because the adoption of multigrade classes is more diffuse at this level of education (quite often small municipalities do not have secondary schools). We restrict our analysis to students in the 5th grade since their potential outcomes include standardized test scores and a number of non-cognitive skills.⁵ Data are from the 2011/12 wave for which we also have information on the actual number of multigrade classes within a school, thanks to additional administrative data (*Rilevazione integrativa, Ministero dell'Istruzione, dell'Università e della Ricerca, MIUR*).

In addition, to reduce the risk of self-selection of students and teachers in different classes and to better identify students effectively attending a multigrade class we restrict to schools that have no more than a classroom for the 5th grade. To compare students attending classes of similar size, we also exclude from our analysis classes with more than 13 students and with less than 5 students (respectively the maximum and the minimum number of students allowed in multigrade classes, applying the discretionary margin of 10%). Finally, we restrict our sample to students that undertook both the literacy and numeracy test.⁶ Our final

⁵ Using data on test scores of 2nd grade students we find that being placed in a multigrade class negatively affects student performance. The effect is smaller in magnitude compared to what we find for 5th graders. For 2nd grade students no information is available on non-cognitive skills (see Tables A8 and A9 in the Appendix of the paper).

⁶ More precisely we consider students for whom there is available both the Rash Literacy Score and the Numeracy Score. As the literacy and numeracy tests were held on different days if we do not restrict the sample to students

sample consists of 14,290 observations in 1,447 schools. This amounts to about 3% of pupils attending the 5th grade in Italian primary schools in 2011-12.

Using the information on the actual number of multigrade classes in the school, provided by administrative sources, we try to infer whether students in our dataset were effectively placed in a multigrade class. Since we do not have detailed information at class level we rely on two different definitions. The first one considers as students attending a multigrade class only those who are enrolled in a school in which there are two multigrade classes.⁷ Following this rule we build a dummy variable *Actual Multigrade1*. As we are focusing on small schools and on small classes, it is very likely that this definition catches students who are effectively attending a multigrade class. According to this measure about 24% of students in our sample are placed in a multigrade class. On the contrary, if we look at the whole sample of students (Table A.1 in the Appendix) we find that about 1% of students are placed in a multigrade class. This percentage seems consistent with the total number of students attending a multigrade class in Italian primary schools. However, we are misplacing those students attending a multigrade class in a school in which the total number of students enrolled is so low that teaching activity is organized in a single multigrade class (covering from grade 1 to grade 5). These students end up included in the control group.

The second definition, *Actual Multigrade2*, is wider and considers as students attending a multigrade class those who attend a school in which there is at least one multigrade class.⁸ Under this second definition, the percentage of students placed in multigrade classes increases to 39%, while if we consider the whole sample we find that about 2% of students are placed in a multigrade class. This definition allows the inclusion of all students attending a multigrade class in the “treated group”, but it is likely that also some students actually not attending a multigrade class end up considered as “treated”.

As regards the outcome variables, the public use files provide two alternative measures of student performance at standardized tests: a) the fractions of correct answers in literacy and numeracy multiple choice tests (*Literacy Score* and *Numeracy Score*); b) scores computed by INVALSI applying the IRT Rasch model to students’ answers in the tests, in order to account for different difficulties of single items (*Rasch Literacy Score* and *Rasch Numeracy Score*)⁹. Since the data come from a national test which is common to all schools, the performance of students attending the same grade are by construction comparable across

undertaking both tests we end up with two slightly different samples for the literacy and numeracy outcomes depending on the number of students that were absent during one of the two tests. Results do not change when using full available samples.

⁷ Since we have already restricted the sample to schools where there is just one 5th grade class, it is rather likely that there must be one class for each grade. Thus having two multigrade classes in the school should imply one grouping 1st and 2nd grades and another with 4th and 5th grades, with 3rd grade being allocated taking into account both class size considerations and similarities between grades curricula. Typically, as curricula taught in 4th and 5th grades are similar (in addition, 5th graders are required to take a final examination at the end of the primary school cycle), in schools having two multigrade classes there is one grouping 1st, 2nd and 3rd grades and another grouping 4th and 5th grades.

⁸ We have also experimented with a very restrictive definition considering as students attending a multigrade class only students who are enrolled in a school in which there is at least a multigrade class and in which the total number of students enrolled in the school divided by the number of multigrade classes in the school is smaller than the maximum of students allowed for multigrade classes. We find qualitatively the same results reported in the paper but the effects are larger in magnitude.

⁹ In such a case these scores are standardized to have a mean of 200 and a standard deviation of 40.

schools in different geographical areas of the country. In order to avoid problems deriving from score manipulation, both these measures are expressed as “cheating-corrected” test scores.¹⁰ However, our main results remain almost identical when using the original scores not corrected for cheating.¹¹

We have also information on the marks assigned by teachers in the two main subjects in the primary school program: Italian Language and Mathematics. Data collected by INVALSI allow the distinction between “written marks” and “oral marks”. We have considered oral marks (*Teacher Marks Numeracy*, *Teacher Marks Literacy*), but results do not change qualitatively if we consider the written marks or the average value of written and oral marks.¹² Teachers’ marks and INVALSI test scores are positively correlated (the correlation between *Rasch Literacy Score* and Italian Language oral mark is 0.51, while the correlation between *Rasch Numeracy Score* and Maths oral mark is 0.38) but there are some relevant differences. The INVALSI tests are identical across schools while marks given by teachers are based on a standard autonomously set by each teacher. Then, while INVALSI scores are comparable across schools, this is not the case for teachers’ marks. In addition, while INVALSI tests assess student performance on an absolute grading scale, teachers might adopt relative marking which might also be affected by class composition.

Apart from measures of cognitive skills, the INVALSI dataset also allows to build some measures of non-cognitive skills. Using the survey submitted to students (Student Questionnaire) the same day of one of the two tests, we consider eight questions allowing us the construction of a locus of control measure, i.e. the extent to which a person believes her\his actions affect her\his outcomes. Five of these questions refer to successful situations¹³ and three of them to unsuccessful ones¹⁴. The student is asked to attribute each situation to: 1) help or lack of help from others; 2) lucky or unlucky circumstances; 3) easiness or difficulty of the task; 4) own ability; 5) own effort. The choice of the first three options denotes an external locus of

¹⁰ As documented by Angrist et al (2017) and Bertoni et al. (2013) many schools follow a “cheating to the test” practice. Since cheating significantly affects the reliability of test scores, INVALSI has developed a statistically solution to purge the data from this problem. This method exploits the statistical properties of the distribution of answers given in classes where the test is taken under the supervision of external examiners (randomly assigned to selected classes and schools with the task of monitoring), and calculates a continuous class-level probability of manipulation (similar to that estimated in Angrist et al. 2017). This probability is based on the variability of intra-class percentage of correct answers, modes of wrong answers, etc.; the resulting estimates are used to “deflate” the raw scores in the test. For a detailed description of the method see INVALSI (2010).

¹¹ These results are not reported for space reasons but are available upon request.

¹² The correlation between written and oral mark is 0.9619 and 0.9712 for Literacy and Numeracy Scores respectively (p-value 0.000).

¹³ This is the list of questions: “1) The teacher asked you to draw an object and you executed it very well. How did you do it?; 2) The teacher asks you to repeat a story you read together in class and you did it very well. How did you do it?; 3) On the first day of school, the teacher asks you to tell what you did during the holidays, you tell it so well that all your schoolmates have fun. How did you do it?; 4) At the recital at the end of the year you performed your part so well that everyone applauded. How did you do it?; 5) The teacher asked you to do a math exercise on the blackboard and you did it very well. How did you do it?” Possible answers are: “I have been helped”, “I have been lucky”, “It was an easy task”, “I am capable” or “I engaged myself”.

¹⁴ This is the list of questions. “1) The teacher asks you to write a composition, but you make many mistakes. How did it happen ? 2) The teacher asks you to repeat a poem you have learned by heart in the past, but you do not remember it very well and make a lot of mistakes. How did it happen ? 3) The teacher asks you to do a work for Christmas, but it comes out very bad and you had to repeat it. How did it happen ?” Possible answers are: “None helped me”, “I have been unlucky”, “It was a hard task”, “I am incapable” or “I did not engage myself”.

control (outcomes depend on luck or external factors), while the choice of the last two options is considered as an indicator of an internal locus of control (outcomes depend on own ability and effort). Then, for each of these questions we build a dummy variable taking the value of 1 when the student picks one of the first three possible answers and zero otherwise. Using students' answers to these questions we build three different measures of locus of control. The first, *External Locus of Control*, is based on the whole set of questions and takes values from 0 (when the student answering to the 8 questions never chooses one of the first three options) to 8 (when the student chooses one of the first three options for each of the 8 questions). The second *External Locus of Control Positive* is based only on questions proposing successful situation and takes values from 0 to 5. The third *External Locus of Control Negative* is instead based exclusively on questions proposing unsuccessful situations and takes values from 0 to 3.

The dataset at hand also provides information on a number of pupils' and parents' characteristics (gender, citizenship, attendance of pre-primary school, parent working status and education). Information on the family background of the student are used by INVALSI to build an indicator of socio-economic status (called ESCS-Economic and Social Cultural Status)¹⁵, out of which principal component analysis is applied, obtaining a variable with zero mean and unitary standard deviation. We also have information on whether the student is younger or older than a regular student, from which we build a dummy variable for students who went to school one year before the suggested age, *Early Enrolled*, and dummy variable for students who entered the school one year after or repeated one or more years, *Late Enrolled*.

We also have information on the number of students enrolled in each grade at the beginning of the school year. For single grade classes the number of enrolled students in the grade corresponds to class size, instead for multigrade classes class size is constructed considering the number of enrolled students in the different grades composing the class.¹⁶

As regards school organization, we know whether the class follows a full-day or half-day schedule and on the basis of this information we build a dummy variable *Full day* for those classes whose schedule is organized in entire days (8am-4pm usually) instead of that only in the morning.

Finally, we have information on the region in which the school is located and on a number of different school catchment area characteristics (population size, extension and altitude).

¹⁵ This indicator is built in accordance to the one proposed in the OECD-PISA framework and considers parents' occupation, educational attainment and possession of educational resources at home (for instance, the number of books). For a detailed description see Ricci (2010), <http://new.sis-statistica.org/wp-content/uploads/2013/09/RS10-SP-The-Economic-Social-and-Cultural-Background-a-continuous-index-for-the-Italian-Students-of-the-fifth-grade.pdf>

¹⁶ To have information on students enrolled in 4th grade in the academic year 2011/12, we consider the information provided by the 2012/13 INVALSI wave when these students are in 5th grade (no test is undertaken in fourth grade). Since, as we explain in more detail below, retention is quite rare in Italian primary schools, this number is likely to be quite close to what it was the year before. Then, for multigrade classes we calculate class size summing the number of students in 5th grade to the number of students enrolled in 4th grade: if this number does not reach the minimum class size imposed for multigrade classes we also add the number of students enrolled in 3rd grade, obtained from the 2013/14 INVALSI wave.

In Panel (a) of Table 1 we report descriptive statistics for the whole sample used in our analysis.¹⁷ *Predicted Multigrade* takes an average value 0.325, implying that according to rules defined by the Italian law about 32% of students in our sample should be placed in a multigrade class. The actual number of students in our sample effectively attending a multigrade class is however smaller and equal to 24% (*Actual Multigrade1*).

Descriptive statistics for the sample of students placed in single grade classes are reported in Panel (b) of Table 1, while in Panel (c) are reported descriptive statistics for students in a multigrade class (*Actual Multigrade1*).¹⁸

If we compare students in the sample used for our main analysis (Table 1, Panel (a)) with the total population of students enrolled in grade 5th (see Table A.1 in the Appendix of the paper) we find a number of differences with respect to individual background that depend on the fact that the students we consider in our analysis live in small villages, typically characterized by poorer economic conditions. In fact, students in our sample are from less wealthy families, less likely to have an immigrant background and to have attended pre-primary school. They obtain worse scores both in literacy and numeracy, while grades assigned by teachers are only slightly lower compared to those observed for the whole population. On the contrary, there are not relevant differences as regards the percentage of students regularly enrolled.

When we compare students in our sample placed in single grade classes (Table 1, Panel (b)) with those placed in multigrade ones (Table 1, Panel (c)) we find that students in multigrade classes are comparable in terms of a number of observable characteristics, such as gender, ESCS index, father and mother nationality. There are, however, some statistically significant differences in terms of percentage of regular students, class size, and attendance of pre-primary school. In addition, the average score obtained both in literacy and numeracy standardized tests is lower for students in multigrade classes compared to students in single grade classes.

¹⁷In the Appendix of the paper (Table A.2) we report descriptive statistics for the whole sample of 5th grade students undertaking the INVALSI standardized test and attending schools that have no more than a classroom in the 5th grade.

¹⁸ In the Appendix of the paper, Table A.3, we report descriptive statistics using the alternative measure *Actual Multigrade2*.

Table1. Descriptive statistics – schools with no more than one classroom for the 5th grade

	Panel (a) Whole sample			Panel (b) Single Grade			Panel (c) Actual Multigrade1		
	Mean	St. Dev.	Obs.	Mean	St. Dev.	Obs.	Mean	St. Dev.	Obs.
<i>Actual Multigrade1</i>	0.241	0.427	14,290						
<i>Actual Multigrade2</i>	0.387	0.487	14,290						
Predicted Multigrade	0.325	0.468	14,290	0.198	0.399	10,852	0.726	0.446	3,438
Rasch Literacy Score	197.224	41.523	14,290	197.990	41.744	10,852	194.808	40.729	3,438
Literacy Score	75.603	13.945	14,290	75.844	14.012	10,852	74.846	13.707	3,438
Teacher Mark Literacy	7.487	1.109	13,200	7.489	1.115	9,979	7.481	1.089	3,221
Rasch Numeracy Score	173.735	41.422	14,290	175.314	40.998	10,852	168.753	42.352	3,438
Numeracy Score	52.943	19.604	14,290	53.219	19.658	10,852	52.072	19.411	3,438
Teacher Mark Numeracy	7.577	1.152	13,252	7.579	1.159	10,014	7.571	1.130	3,238
External Locus of Control	2.247	1.667	14,119	2.223	1.667	10,717	2.322	1.663	3,402
Ext Locus Control Positive	1.364	1.162	13,895	1.351	1.164	10,546	1.404	1.156	3,349
Ext. Locus Control Negative	0.912	0.964	14,006	0.901	0.959	10,630	0.947	0.977	3,376
Female	0.491	0.500	14,290	0.491	0.500	10,852	0.492	0.500	3,438
Regularly Enrolled	0.942	0.233	14,290	0.939	0.238	10,852	0.952	0.214	3,438
Early Enrolled	0.015	0.122	14,290	0.017	0.128	10,852	0.010	0.100	3,438
Late Enrolled	0.042	0.202	14,290	0.044	0.205	10,852	0.038	0.191	3,438
Pre Primary School	0.115	0.319	14,290	0.123	0.328	10,852	0.090	0.286	3,438
ESCS index	-0.009	0.973	14,290	0.000	0.987	10,852	-0.037	0.925	3,438
Italian Father	0.850	0.357	14,290	0.847	0.360	10,852	0.860	0.347	3,438
Italian Mother	0.831	0.375	14,290	0.830	0.375	10,852	0.834	0.372	3,438
Full day	0.161	0.367	14,290	0.166	0.372	10,852	0.145	0.352	3,438
# student enrolled in grade 5	10.278	2.548	14,290	11.097	1.892	10,852	7.695	2.627	3,438
Class size	10.555	2.152	14,290	11.101	1.876	10,852	8.829	2.055	3,438
Southern regions	0.371	0.483	14,290	0.380	0.485	10,852	0.341	0.474	3,438
Altitude	382.074	293.694	14,290	343.794	282.104	10,852	502.905	296.851	3,438
Population size	45,966.97	250,921.2	14,290	58,477.34	286,547.3	10,852	6,478.158	21,765.03	3,438

Notes: The data are obtained from the Invalsi websites (downloaded on 7/2/2016)

4. Estimation Strategy

Distinguishing the effect of grade mixing from the effect produced by other factors that are both relevant for student cognitive and non-cognitive skills and for being placed in a multigrade class is not an easy task. In order to recover the causal impact of multigrade classes on student performance we exploit the rule that sets to 10 the minimum number of students in a class and also requires that multigrade classes cannot be composed by more than 12 students. As these rules introduce a plausibly exogenous variation in treatment status, we use them to identify the impact of multigrade classes on student outcomes. In fact, the application of these rules implies that the probability of attending a multigrade class is a discontinuous function of the number of students enrolled in a class and also depends on the number of students enrolled in adjacent grades. Then, we apply an instrumental variable strategy that uses the class size rule as an instrument for being placed in a multigrade class. We estimate the following model:

$$[1] \quad Y_{ij} = \beta_0 + \beta_1 \text{Actual Multigrade}_{ij} + \beta_2 \text{Class size}_j + \beta_3 X_{ij} + \lambda_k + \epsilon_{ij}$$

$$[2] \quad \text{Actual Multigrade}_{ij} = \phi_0 + \phi_1 \text{Predicted Multigrade}_{ij} + \phi_2 \text{Class size}_j + \phi_3 X_{ij} + \lambda_k + v_{ij}$$

where in equation [1] Y_{ij} is the outcome variable of interest (alternatively the performance of student i enrolled in class j in Literacy and Numeracy or her\his indicator of locus of control); $\text{Actual Multigrade}_{ij}$ is a dummy variable indicating whether the student is attending a multigrade class; Class size_j is the number of students attending class j ; X_{ij} is a vector of individual and school characteristics (gender, ESCS index, mother and father immigrant status, pre-primary school attendance, Early Enrolled, Late Enrolled, Full day); λ_k are regional fixed effects and ϵ_{ij} is a random error term. Equation [2] represents the first stage of the relationship between student actual placement in a multigrade class and $\text{Predicted Multigrade}_{ij}$, that is a dummy variable taking the value of one when the number of students enrolled in 5th grade class j attended by student i is smaller than 10 and when class size in a multigrade class does not exceed the maximum number of 12 students allowed by the law.

The rule establishing a specific maximum number of students for multigrade classes allows us to avoid problems typically encountered by studies that exploit minimum class size rules. These works have to take into account that when a multigrade class is formed there is also a sharp change in class size. In our case, since according to the Italian law, multigrade classes cannot be composed by more than 12 students, class size is similar in multigrade and single grade classes that are just above the minimum threshold of 10 students. In our sample, average class size in multigrade classes is of 8.8, while it is equal to 11 students in single grade classes. However, as described in equation [1], in our estimates we control for class size: to avoid problems that might derive from class size endogeneity we have restricted our analysis to small schools that have no more than one classroom for 5th grade and to classes composed by no more than 13 students and no less than 5 students (the maximum and minimum class size imposed by the reform for multigrade classes, applying the discretionary margin of 10%).

Given the type of schools considered in our analysis is also unlikely that school managers behaved in such a way as to change the number of enrolled students in a given grade when it was near the cut-off point (the composition of a multigrade class also depends on the number of students enrolled in adjacent grades). They might be interested in such manipulations to avoid the reduction of the number of teachers working in the school or (at the opposite extreme) to get rid of undesirable teachers. However, the minimum class size rule applies when the number of enrolled students becomes quite low and this is likely to occur in small and/or isolated municipalities where it is difficult to attract new students. In addition, in Italy grade retention in primary school is very rare and it is very unlikely that teachers and school managers use this practice to reach the minimum class size rule. In our sample 94% of students are regular in their school path, while about 4.2% and 1.5% of them are late or early entrants in their educational path. Delays in students' educational process are mostly due to the fact that non-Italian students are often placed in grades lower than those corresponding to their age at the entry in the country, in order to improve the mastery of Italian

language. In fact, when we only consider native students who typically start school at the expected age, the percentage of students being late in their educational career drops to 2%.

In order to try to understand whether there has been manipulation in the number of students enrolled in 4th and 5th grades, in Figure 2 and Figure 3 we present the density of grade enrolment in each grade for our sample and the whole sample (including also classes with more than 13 students), respectively. As shown in Figure 2 that plots the density of grade enrolment, there are no suspect discontinuities.

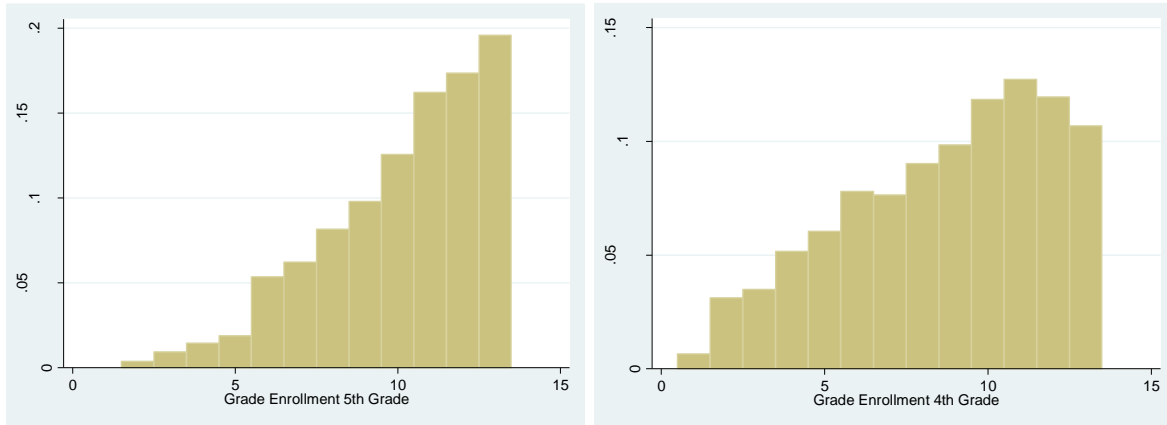


Figure 2. Density of Grade Enrolment, small classes sample

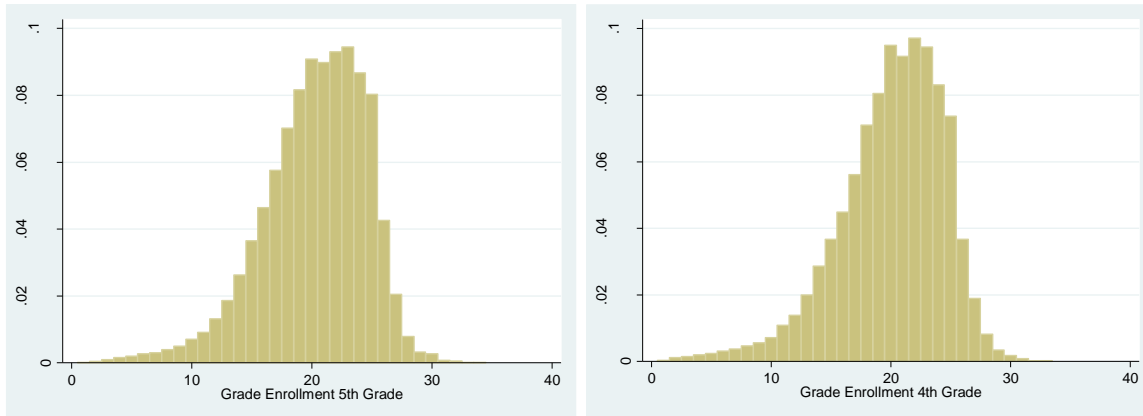


Figure 3. Density of Grade Enrolment, full sample

We have also checked whether our instrument is correlated to student and school characteristics. Even if testing exclusion restriction is not possible since it involves the structural error that is never observable, if there is no relationship between the instrument and observable student and school characteristics one may reasonably argue that also unobserved student and school characteristics be uncorrelated with the instrument. Similarly, to Sims (2008) and Sattari (2016), we regress students' and schools' characteristics on Predicted Multigrade_{ij}. Results of our regression are reported in Table 2, where we test whether the Predicted Multigrade_{ij} is predictive of a number of student characteristics. Overall, Table 2 confirms that

these variables typically fail to show a statistically significant correlation with instrument status. However, we control for these variables in the regressions to avoid any bias due to the lack of balance.¹⁹

Table2. Differences in predetermined characteristics. 5th grade – Italy 2011-12

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)
	Female	Regularly Enrolled	Early Enrolled	Late Enrolled	Pre- primary school	ESCS index	Italian Father	Italian Mother	Full Day	Altitude	Popul. size
Predicted Multigrade	0.007 (0.008)	0.010** (0.005)	-0.006*** (0.002)	-0.006 (0.004)	-0.015* (0.008)	-0.033 (0.024)	-0.001 (0.009)	-0.008 (0.009)	-0.021 (0.019)	143.066*** (14.453)	-51.999*** (11.613)
Obs	14,290	14,290	14,290	14,290	14,290	14,290	14,290	14,290	14,290	14,290	14,290

The symbols ***, **, * indicate that the coefficients are statistically significant at the 1, 5 and 10 percent level, respectively.

As explained in the previous section, we observe an imperfect correspondence with the Assigned Treatment ($\text{PredictedMultigrade}_{ij}$) and Actual Multigrade. This is due both to the fact that effective treatment is measured with error and to the fact that schools may deviate from the rule. Based on the First Stage Equation, in Figures 4, we plot the probability of attending a multigrade class against class size when considering the whole sample of students attending the 5th grade in schools with no more than 5 classrooms (estimates for this sample are reported in the Appendix of the paper). The circles are the means of the probability of effectively attending a multigrade class for a given Class Size, while the red dots are the predicted values from the first stage equation. As it is possible to see on the left hand side of the graph in Figure 4, the probability of effectively attending a multigrade class for students in classes just below the cut-off point is about 0.39 while it drops 0.19 for students in classes with a number of enrolled students just above the threshold. A similar picture emerges also when focusing on the sample we use in our main analysis (Figure 5).

¹⁹ In the Appendix of the paper we report a number of figures aimed at analyzing whether our control variable show relevant discontinuities at the cutoff point of 10 students. In these figures we present a number of school and student characteristics plotted against the number of students enrolled in 5th grade.

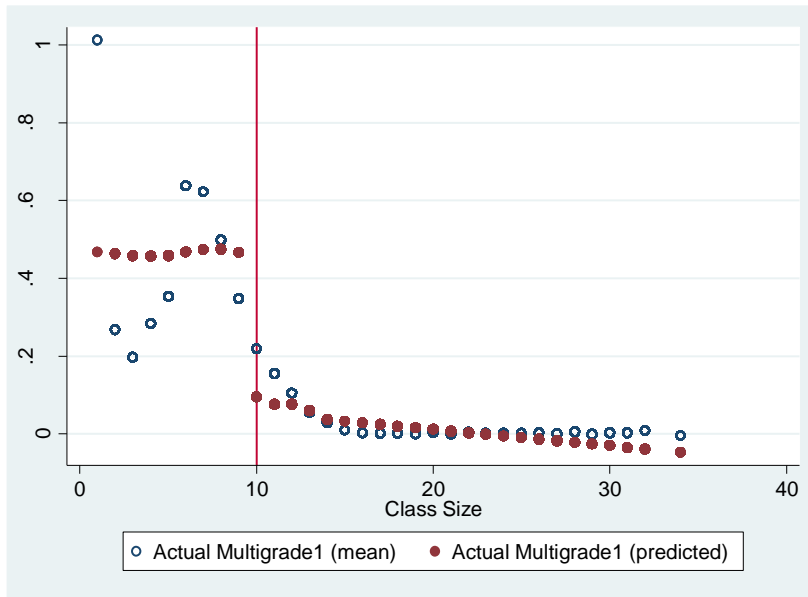


Figure4. First Stage relationship: Grade Enrolment and Predicted Actual Multigrade. Full sample

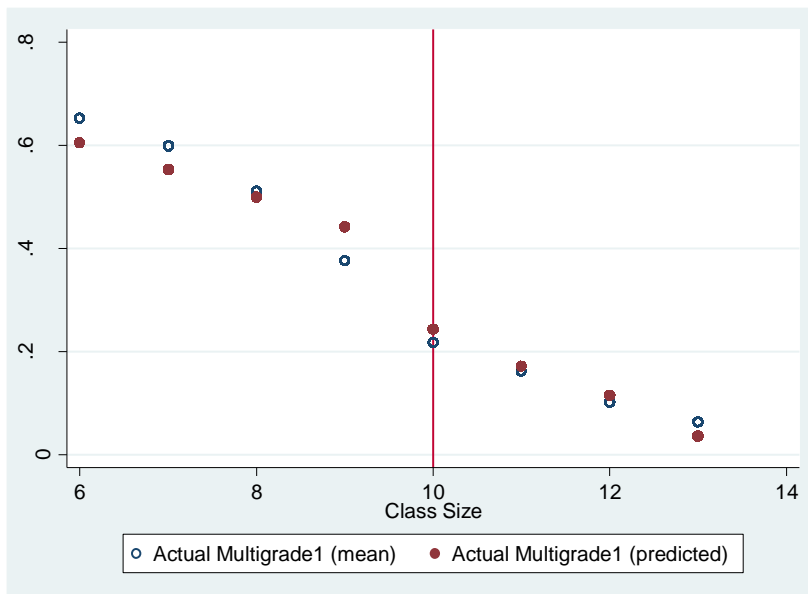


Figure 5. First Stage relationship: Grade Enrolment and Predicted Actual Multigrade. Small classes sample

5. The effect of multigrade classes on cognitive skills

In this section, we report our main results of the effect of being placed in a multigrade class on some measures of student cognitive abilities. The INVALSI dataset provides information on both standardized tests and teacher evaluations. Then, for each student we observe both the marks assigned by math and language teachers and the INVALSI standardized test score results obtained in the same areas during the

same school year. We firstly analyse the impact of multigrade classes on the INVALSI standardized test scores and then we focus our attention on marks assigned by teachers.

5.1. The effect of multigrade classes on standardized test scores

We begin our analysis using students' performance in standardized test scores as outcome variable. Initially we focus on student performance in literacy test and then we turn our attention to their performance in numeracy.

In Table 3 we report instrumental variable estimation results when considering as dependent variable the Literacy Score, while showing in Panel B of the same Table First Stage estimation results. Standard errors are robust to heteroscedasticity and are allowed for clustering at the class level. In all specifications we control for regional fixed effects.

The First Stage shows that Predicted Multigrade strongly determines the effective treatment, since the First Stage F-statistics is always greater than 57.

In the first two specifications of Table 3 our dependent variable is the fraction of correct answers in the Literacy test. In specification (1) we control for school and individual characteristics. We find that being placed in a multigrade class produces a negative effect on student performance in the Literacy test. The effect remains statistically significant (at 10 percent level) when we control for class size (column 2). Students in multigrade classes obtain a lower *Literacy Scores* by about 0.2 a standard deviation when compared to students in single grade classes. Similar results are found in specifications (3) and (4) where we replicate the first two specifications of Table 3 but consider as outcome variable the Rasch score, which also take into account the different degree of difficulty of questions. However, the effect is not statistically significant at conventional levels when we control for class size. Class size does not produce any statistically significant effect.

The effects of control variables are consistent with the findings presented in the existing literature. Females tend to perform better than males. Students with a better socio-economic background obtain better results compared to students who are from more disadvantaged families. In addition, students with Italian parents perform better than students whose parents were born abroad.²⁰

²⁰As Early Enrolled, Late Enrolled and Full Day might be bad controls we have also estimated all our models without these control variables. Results remain qualitatively unchanged.

Table 3. TSLS Estimates of Multigrade Classes on Literacy Scores

	(1)	(2)	(3)	(4)
	<i>Literacy Score</i>	<i>Literacy Score</i>	<i>Rasch Literacy Score</i>	<i>Rasch Literacy Score</i>
Panel A: Two-Stage Least Squares Estimates				
<i>Actual Multigrade1</i>	-2.849*** (0.990)	-3.981* (2.341)	-8.211*** (2.795)	-9.642 (6.786)
<i>Class Size</i>		-0.117 (0.209)		-0.148 (0.602)
<i>Female</i>	2.674*** (0.231)	2.675*** (0.231)	7.779*** (0.684)	7.779*** (0.684)
<i>Pre Primary School</i>	-0.186 (0.401)	-0.211 (0.407)	-0.982 (1.189)	-1.014 (1.199)
<i>Early Enrolled</i>	-1.785* (1.053)	-1.816* (1.055)	-5.622* (2.961)	-5.660* (2.958)
<i>Late Enrolled</i>	-6.651*** (0.742)	-6.653*** (0.741)	-17.092*** (1.921)	-17.094*** (1.921)
<i>Full day</i>	-0.398 (0.568)	-0.399 (0.570)	-1.072 (1.656)	-1.073 (1.659)
<i>ESCS index</i>	2.839*** (0.142)	2.829*** (0.144)	8.938*** (0.406)	8.926*** (0.411)
<i>Italian Father</i>	2.174*** (0.445)	2.193*** (0.446)	6.476*** (1.297)	6.501*** (1.300)
<i>Italian Mother</i>	1.320*** (0.398)	1.323*** (0.398)	4.365*** (1.177)	4.369*** (1.178)
<i>Altitude</i>	0.001 (0.001)	0.001 (0.001)	0.003 (0.003)	0.003 (0.003)
<i>Population</i>	-0.000 (0.001)	-0.000 (0.001)	0.000 (0.002)	0.000 (0.002)
Observations	14290	14290	14290	14290
Panel B: First Stage				
<i>Predicted Multigrade</i>	0.392*** (0.007)	0.288*** (0.012)	0.392*** (0.007)	0.288*** (0.012)
<i>Class Size</i>		-0.027*** (0.003)		-0.027*** (0.003)
First Stage F Statistics	318.651	57.448	318.651	57.448

Notes: The Table reports TSLS estimates. Standard errors in parentheses are corrected for heteroscedasticity. In all regressions we control for regional dummies (20 categories, not reported). The symbols ***, **, * indicate that coefficients are statistically significant, respectively, at the 1, 5, and 10 percent level.

In Table 4 we report results obtained when considering as dependent variable student cognitive skill in numeracy. Similarly to Table 3, in the two first specification of Table 4 we consider as outcome variable the number of correct answers in numeracy test, while specifications (3) and (4) consider the *Rash Numeracy Score*. We find that being placed in a multigrade class reduces student score in numeracy. The effect remains also when controlling for class size: students in multigrade classes obtain a lower *Numeracy Scores* by about half standard deviation when compared to students in single grade classes. This is a significant magnitude: a decrease of half standard deviation implies moving from the mean (which is also about the median of the distribution) down to the third decile. Class size does not produce any statistically significant impact.

The negative effect holds true when we consider as outcome variables the Rasch score. The magnitude of the effect is of about 0.4 of a standard deviation. In order to evaluate the magnitude of the estimated effect, it is useful to consider that the effect of being assigned to a multigrade class on the *Rasch Numeracy Score* corresponds to the effect produced by an increase in *ESCS index* of about 3 standard deviations.

Table 4. TSLS Estimates of Multigrade Classes on Numeracy Scores

	(1)	(2)	(3)	(4)
	<i>Numeracy Score</i>	<i>Numeracy Score</i>	<i>Rasch Numeracy Score</i>	<i>Rasch Numeracy Score</i>
Panel A: Two-Stage Least Squares Estimates				
<i>Actual Multigrade1</i>	-5.448*** (1.569)	-9.277** (3.954)	-14.710*** (3.235)	-12.611 (7.885)
<i>Class Size</i>		-0.396 (0.336)		0.217 (0.679)
<i>Female</i>	2.215*** (0.330)	2.214*** (0.332)	2.705*** (0.584)	2.705*** (0.583)
<i>Pre Primary School</i>	0.128 (0.596)	0.041 (0.611)	3.023** (1.310)	3.070** (1.320)
<i>Early Enrolled</i>	-0.180 (1.565)	-0.283 (1.564)	6.938** (3.196)	6.994** (3.194)
<i>Late Enrolled</i>	-6.602*** (0.871)	-6.608*** (0.874)	-7.029*** (1.653)	-7.026*** (1.649)
<i>Full day</i>	0.803 (0.818)	0.801 (0.834)	-1.199 (1.711)	-1.198 (1.708)
<i>ESCS index</i>	4.075*** (0.201)	4.042*** (0.207)	5.048*** (0.421)	5.066*** (0.429)
<i>Italian Father</i>	2.081*** (0.649)	2.147*** (0.657)	2.842** (1.195)	2.806** (1.193)
<i>Italian Mother</i>	1.561*** (0.584)	1.571*** (0.590)	1.365 (1.108)	1.360 (1.105)
<i>Altitude</i>	0.005*** (0.001)	0.005*** (0.002)	-0.001 (0.003)	-0.002 (0.004)
<i>Population</i>	-0.000 (0.002)	-0.001 (0.002)	0.010*** (0.003)	0.010*** (0.003)
Observations	14290	14290	14290	14290
Panel B: First Stage				
<i>Predicted Multigrade</i>	0.392*** (0.007)	0.288*** (0.012)	0.392*** (0.007)	0.288*** (0.012)
<i>Class Size</i>		-0.028*** (0.003)		-0.028*** (0.003)
First Stage F Statistics	318.651	57.448	318.651	57.448

Notes: The Table reports TSLS estimates. Standard errors (reported in parentheses) are corrected for heteroscedasticity. In all the regressions we control for regional dummies (20 categories, not reported). The symbols ***, **, * indicate that coefficients are statistically significant, respectively, at the 1, 5, and 10 percent level.

Comparing results presented in Tables 3 and 4 with those obtained with an OLS estimator (see Table A.4 in the Appendix of the paper), we find that even if OLS and IV estimates yield the same qualitatively results the magnitude of the effects is higher in IV estimates pointing to a bias in OLS estimates possibly due to omitted variables or measurement error.

Qualitatively the same results are found, both for the Literacy and Numeracy test scores, when we run our regressions considering the whole sample of students undertaking the INVALSI standardized test and attending schools with no more than five classrooms (see Table A.4. in the Appendix of the paper). In addition, these results are robust when controlling for a number of characteristics of the municipality in which the school is located, such as population, extension in squared kilometres, altitude (results not reported but available upon request).²¹ We also checked whether the effect of multigrade classes is heterogeneous

²¹ During the day of the INVALSI test 4th grade students are placed in a different classroom and engaged with standard class activities. This implies that 5th grade students in multigrade classes should not suffer the negative spillovers deriving from an environment where some students are not concentrated on the same test.

according to student gender and economic background, but we do not find statistically significant differences.

5.2. The effect of multigrade classes on marks assigned by teachers

In this section we investigate the effect of multigrade classes on student performance as assessed by the teachers through the marks assigned on the subject. In Table 5 we replicate the same specifications we have estimated in the previous analysis but considering as outcome variables the marks assigned by teachers in Literacy (columns 1 and 2) and Numeracy (columns 3 and 4), respectively. In specifications where we do not control for class size, we do not find any statistically significant effect of being assigned to a multigrade class on teacher assessment of students' skills. When we control for class size we find a negative effect statistically significant at the 10 percent level.

Table 5. TSLS Estimates of Multigrade Classes on marks assigned by teachers

	(1)	(2)	(3)	(4)
	<i>Teacher Mark Literacy</i>	<i>Teacher Mark Literacy</i>	<i>Teacher Mark Numeracy</i>	<i>Teacher Mark Numeracy</i>
Panel A: Two-Stage Least Squares Estimates				
<i>Actual Multigrade1</i>	-0.013 (0.076)	-0.344* (0.178)	0.056 (0.078)	-0.346* (0.184)
<i>Class Size</i>		-0.034** (0.015)		-0.041*** (0.015)
<i>Female</i>	-0.247*** (0.019)	-0.247*** (0.020)	-0.062*** (0.020)	-0.062*** (0.020)
<i>Pre Primary School</i>	-0.035 (0.030)	-0.044 (0.031)	-0.031 (0.031)	-0.043 (0.032)
<i>Early Enrolled</i>	0.089 (0.081)	0.082 (0.081)	0.063 (0.075)	0.055 (0.075)
<i>Late Enrolled</i>	-0.533*** (0.051)	-0.531*** (0.051)	-0.459*** (0.055)	-0.457*** (0.055)
<i>Full day</i>	-0.058 (0.041)	-0.057 (0.041)	-0.033 (0.040)	-0.031 (0.040)
<i>ESCS index</i>	0.304*** (0.011)	0.301*** (0.011)	0.312*** (0.011)	0.309*** (0.011)
<i>Italian Father</i>	0.148*** (0.037)	0.154*** (0.037)	0.112*** (0.039)	0.120*** (0.039)
<i>Italian Mother</i>	0.228*** (0.034)	0.227*** (0.034)	0.216*** (0.035)	0.214*** (0.035)
<i>Altitude</i>	0.000 (0.000)	0.000 (0.000)	0.000 (0.000)	0.000 (0.000)
<i>Population</i>	0.000 (0.000)	0.000 (0.000)	0.000* (0.000)	0.000* (0.000)
Observations	13200	13200	13252	13252
Panel B: First Stage				
<i>Predicted Multigrade</i>	0.408*** (0.022)	0.289*** (0.012)	0.392*** (0.007)	0.289*** (0.012)
<i>Class Size</i>		-0.027*** (0.008)		-0.027*** (0.008)
First Stage F Statistics	292.38	52.259	292.38	52.259

Notes: The Table reports TSLS estimates. Standard errors (reported in parentheses) are corrected for heteroscedasticity. In all the regressions we control for regional dummies (20 categories, not reported). The symbols ***, **, * indicate that coefficients are statistically significant, respectively, at the 1, 5, and 10 percent level.

All in all, we find that students attending multigrade classes perform significantly worse in Numeracy standardized tests, while the effect on Literacy though negative is smaller and in some cases statistically insignificant. On the other hand, the effect of being placed in a multigrade class on teachers' assessment of student skills is negative but not always statistically significant.

6. The effect of multigrade classes on non-cognitive skills

Individual success is both determined by cognitive and non-cognitive abilities (or personality traits). Locus of control has received particular attention by both psychologists and economists who have shown its relevance for social and economic success.

In order to assess the reliability of our measures of external locus of control we have tried to see whether they behave similarly to what described in the literature. One well documented result is the negative correlation between cognitive ability and external locus of control (see Cebi, 2007; Baron and Cobb-Clark, 2010). This relationship is confirmed by our data as we find a negative correlation between student performance in literacy and numeracy and our measures of external locus of control. The correlation between the Numeracy Score and External Locus of Control is -0.167 (p-value 0.000), while the correlation between the Literacy Score and External Locus of Control is -0.199 (p-value 0.000).

In Table 6 we report instrumental variable estimation results when considering different measures of external locus of control as dependent variable. In the first two specifications our dependent variable is External Locus of Control. We find that students attending a Multigrade class are more inclined to have an external locus of control (Table 6, column 1). However, the effect is statistically not significant when we control for class size (column 2). In columns (3) and (4) we consider as dependent variable students tendency to attribute unsuccessful situations to external factors (External Neg). We find that students attending a multigrade class are more inclined to attribute unsuccessful situations to luck or other factors behind their own control (column 3), even when controlling for class size (column 4). On the other hand we do not find any statistically significant correlation between Actual Multigrade and student tendency to attribute successful situations to external factors (columns 5 and 6).

The sign of the effects remains the same also when considering the whole sample of students but the effects become more imprecisely estimated and statistically not significant at conventional levels.

The effect we find can derive both from teachers adopting a different teaching approach and from students interacting in a very peculiar environment. For instance, it could be that multigrade classes involve a more individualized teaching style that might affect socio-emotional skills. In addition, since multigrade classes are more heterogeneous in terms of pupils' age, it might be more difficult for students to understand whether the results they get are due to ability and effort or are instead related to age. If success or un-success depends on being the youngest or the oldest in a group, why bothering in exerting effort? Age heterogeneity can also affect the quality of relationships among peers and through this channel affect locus of control and

so on. Unfortunately, lack of appropriate data (like questions on motivating individual behaviour) does not allow us to further investigate these channels.

Table 6. TSLS Estimates of Multigrade Classes on External Locus of Control

	(1)	(2)	(3)	(4)	(5)	(6)
	<i>External Locus of Control</i>	<i>External Locus of Control</i>	<i>External Locus of Control Neg</i>	<i>External Locus of Control Neg</i>	<i>External Locus of Control Pos</i>	<i>External Locus of Control Pos</i>
Panel A: Two-Stage Least Squares Estimates						
<i>Actual Multigrade1</i>	0.107* (0.093)	0.349 (0.225)	0.090* (0.053)	0.224* (0.136)	0.025 (0.062)	0.122 (0.150)
<i>Class Size</i>		0.025 (0.020)		0.014 (0.012)		0.010 (0.013)
<i>Female</i>	0.311*** (0.028)	0.311*** (0.028)	0.155*** (0.017)	0.155*** (0.017)	0.161*** (0.020)	0.161*** (0.020)
<i>Pre Primary School</i>	0.047 (0.046)	0.052 (0.046)	0.009 (0.028)	0.011 (0.028)	0.047 (0.031)	0.049 (0.031)
<i>Early Enrolled</i>	0.156 (0.138)	0.162 (0.138)	0.011 (0.080)	0.015 (0.081)	0.125 (0.090)	0.127 (0.090)
<i>Late Enrolled</i>	0.410*** (0.076)	0.410*** (0.076)	0.234*** (0.046)	0.235*** (0.046)	0.195*** (0.055)	0.195*** (0.055)
<i>Full day</i>	0.045 (0.052)	0.045 (0.052)	0.026 (0.029)	0.026 (0.030)	0.022 (0.033)	0.022 (0.033)
<i>ESCS index</i>	-0.169*** (0.016)	-0.167*** (0.016)	-0.050*** (0.009)	-0.049*** (0.009)	-0.125*** (0.011)	-0.124*** (0.011)
<i>Italian Father</i>	-0.083 (0.056)	-0.088 (0.056)	-0.073** (0.032)	-0.076** (0.032)	-0.007 (0.039)	-0.008 (0.039)
<i>Italian Mother</i>	-0.112** (0.050)	-0.113** (0.050)	-0.047 (0.029)	-0.047 (0.029)	-0.068* (0.035)	-0.068* (0.035)
	0.000 (0.000)	0.000 (0.000)	-0.000 (0.000)	-0.000 (0.000)	0.000 (0.000)	0.000 (0.000)
	-0.000 (0.000)	-0.000 (0.000)	-0.000 (0.000)	-0.000 (0.000)	0.000 (0.000)	0.000 (0.000)
Observations	14119	14119	14006	14006	13895	13895
Panel B: First Stage						
Predicted Multigrade	0.394*** (0.007)	0.288*** (0.012)	0.394*** (0.007)	0.287*** (0.012)	0.393*** (0.007)	0.287*** (0.012)
Class Size		-0.028*** (0.003)		-0.028*** (0.003)		-0.028*** (0.003)
First Stage F Statistics	318.789	56.946	317.119	56.510	316.03	56.357

Notes: The Table reports TSLS estimates. Standard errors (reported in parentheses) are corrected for heteroscedasticity. In all the regressions we control for regional dummies (20 categories, not reported). The symbols ***, **, * indicate that coefficients are statistically significant, respectively, at the 1, 5, and 10 percent level.

7. Robustness Checks

In this section we perform a number of robustness checks. First, even if we dealt cheating problems by using the “cheating-corrected” test scores, to check whether our results are driven by different cheating behaviour in single and multigrade classes, we restrict to schools located in the Northern regions of the country where cheating problems have been proved to be less important (see for instance Paccagnella and Sestito, 2014). Second, to check whether our results could have been driven by the selection of students with poorer

background in multigrade classes, we restrict our analysis respectively to school located in mountainous areas and to municipalities with not more than one school where it is presumably costly to enrol students in a school located in a different area. Third, to this point our measure of Actual Multigrade has consisted of students attending small classes in small schools where there were at least two multigrade classrooms, we examine the robustness of our results to an alternative measure that includes students enrolled in schools where there is only one multigrade class. Finally, we use an alternative sample selection procedure focusing on small municipalities and instrument class size with population size.

In Table 7 we report estimation results for the specifications with the full set of regressors while restricting the sample to schools located in Northern regions and considering as outcome variables alternatively the Literacy and Numeracy scores (columns 1 and 2), the marks assigned by teachers (columns 3 and 4) and our indicator of external locus of control (column 5).

As regard to student performance on Literacy and Numeracy test scores we find results that are consistent with those found with the sample extended to the full country (we find very similar results also when considering as outcome variables *Rasch Literacy Score* and *Rasch Numeracy Score*, not reported). We also find that Actual Multigrade produces a negative and statistically significant effect on student performance when considering marks assigned by teachers as a measure of their cognitive skills (the effect is negative but far from being statistically significant when we restrict the sample to schools located in the South (see Dardanoni et al., 2009).

Results are also consistent with the previous section when we consider External Locus of Control as outcome variable (column 5): once again we find that being placed in a multigrade class increases students' inclination to attribute success and failure to external factors.

Table 7. TSLS Estimates of Multigrade Classes on test scores, teachers' marks and locus of control. Schools located in Northern regions

	(1)	(2)	(3)	(4)	(5)
	<i>Literacy Score</i>	<i>Numeracy Score</i>	<i>Teacher Mark Literacy</i>	<i>Teacher Mark Numeracy</i>	<i>External Locus Control</i>
Panel A: Two-Stage Least Squares Estimates					
<i>Actual Multigrade</i>	-5.067*	-9.929*	-0.529**	-0.617**	0.516*
	(2.851)	(5.074)	(0.266)	(0.278)	(0.305)
<i>Class Size</i>	-0.045	-0.464	-0.051**	-0.064***	0.040
	(0.240)	(0.420)	(0.022)	(0.023)	(0.026)
Observations	8993	8993	8382	8416	8889
Panel B: First Stage					
Predicted Multigrade	0.261***	0.261***	0.247***	0.247***	0.262***
	(0.016)	(0.016)	(0.016)	(0.016)	(0.016)
Class Size	-0.031***	-0.031***	-0.033***	-0.033***	-0.031***
	(0.003)	(0.003)	(0.003)	(0.003)	(0.003)
First Stage F Statistics	28.172	28.172	22.788	22.814	28.444

Notes: The Table reports TSLS estimates. Standard errors (reported in parentheses) are corrected for heteroscedasticity. In all the regressions we control for our full set of controls and for regional dummies (20 categories). The symbols ***, **, * indicate that coefficients are statistically significant, respectively, at the 1, 5, and 10 percent level.

One might argue that our results are driven by the selection of students with poorer socio-economic background in multigrade classes: while well off parents may have the possibility to enrol their children in

schools that offer single grade classes, less advantaged families tend to avoid the commuting costs and let their children to attend the multigrade class offered by the local school. To deal with this issue we have firstly restricted our sample to schools located in mountainous areas²² where commuting costs are higher. In this sample the difference in ESCS between students attending multigrade and single grade classes is very small (-0.02) and far from being statistically significant (p-value 0.412). As shown in Table 8, our main results remain qualitatively unchanged. As a further check we have analyzed how our estimates react to changes in the measurement error of student socio-economic background. At this aim instead of controlling for the synthetic measure of socioeconomic status (ESCS), we have used alternative measures of students' background separately and different combinations of them (the number of years of education of fathers and mothers, four dummy variables for the number of books at home, two dummy variables for Father Unemployed and Mother Unemployed). Regardless of the measures of socio economic condition used, students in multigrade classes show a worse performance compared to students in single grade classes, the coefficient being negative and quite stable in magnitude (results not reported and available upon request).

Table 8. TSLS Estimates of Multigrade Classes on test scores, teachers' marks and locus of control. Schools located in mountainous areas

	(1)	(2)	(3)	(4)	(5)
	<i>Literacy Score</i>	<i>Numeracy Score</i>	<i>Teacher Mark Literacy</i>	<i>Teacher Mark Numeracy</i>	<i>External Locus Control</i>
Panel A: Two-Stage Least Squares Estimates					
<i>Actual Multigrade1</i>	-6.007** (2.786)	-10.796** (4.626)	-0.370* (0.206)	-0.490** (0.220)	0.343 (0.265)
<i>Class Size</i>	-0.332 (0.266)	-0.403 (0.418)	-0.036* (0.019)	-0.054*** (0.020)	0.018 (0.025)
Observations	9181	9181	8606	8614	9088
Panel B: First Stage					
Predicted Multigrade	0.295*** (0.015)	0.295*** (0.015)	0.299*** (0.016)	0.299*** (0.016)	0.296*** (0.015)
Class Size	-0.032*** (0.003)	-0.032*** (0.003)	-0.032*** (0.003)	-0.032*** (0.003)	-0.032*** (0.003)
First Stage F Statistics	38.207	38.207	36.463	36.632	38.321

Notes: The Table reports TSLS estimates. Standard errors (reported in parentheses) are corrected for heteroscedasticity. In all the regressions we control for our full set of controls and for regional dummies (20 categories). The symbols ***, **, * indicate that coefficients are statistically significant, respectively, at the 1, 5, and 10 percent level.

As an additional robustness check we have also restricted the sample to municipalities with not more than one school, implying that parents who want to move their child to another school should also change municipality suffering additional costs. As shown in Table 9 results remain substantially unchanged.

²² Mountainous areas are defined by law 991/1952 referring both to altitude and economic conditions (see <http://www.gazzettaufficiale.it/eli/id/1952/07/31/052U0991/sg>). Comparing the altitude of mountainous areas with that of non-mountainous areas we find that the average altitude is of 504 and 163 meters respectively.

Table 9. TSLS Estimates of Multigrade Classes on test scores, teachers' marks and locus of control. Only municipalities with not more than one school

	(1)	(2)	(3)	(4)	(5)
	<i>Literacy Score</i>	<i>Numeracy Score</i>	<i>Teacher Mark Literacy</i>	<i>Teacher Mark Numeracy</i>	<i>External Locus Control</i>
Panel A: Two-Stage Least Squares Estimates					
<i>Actual Multigrade1</i>	-4.166*	-8.835**	-0.344*	-0.348*	0.358
	(2.375)	(3.978)	(0.183)	(0.188)	(0.231)
<i>Class Size</i>	-0.093	-0.288	-0.031**	-0.040**	0.027
	(0.213)	(0.340)	(0.016)	(0.016)	(0.021)
Observations	12886	12886	11920	11971	12733
Panel B: First Stage					
Predicted Multigrade	0.293***	0.293***	0.289***	0.291***	0.292***
	(0.013)	(0.013)	(0.013)	(0.013)	(0.013)
Class Size	-0.027***	-0.027***	-0.027***	-0.027***	-0.027***
	(0.003)	(0.003)	(0.003)	(0.003)	(0.003)
First Stage F Statistics	53.792	53.792	47.643	48.484	52.968

Notes: The Table reports TSLS estimates. Standard errors (reported in parentheses) are corrected for heteroscedasticity. In all the regressions we control for our full set of controls and for regional dummies (20 categories). The symbols ***, **, * indicate that coefficients are statistically significant, respectively, at the 1, 5, and 10 percent level.

We now turn to the robustness of our results to an alternative definition of Actual Multigrade. As explained in Section 2, ActualMultigrade1 misplaces some 5th graders in multigrade classrooms in the control group. If the effect of being in a multigrade classroom is negative, such specification underestimates the negative effect. In Table 10 we report results obtained when using as indicator of students being actually placed in a multigrade class the dummy variable Actual Multigrade2. We find qualitatively similar results even if the magnitude of the effects is somehow larger pointing to the fact that when using Actual Multigrade1 some students actually attending a multigrade class were attributed to the control group.²³

Table 10. TSLS Estimates of Multigrade Classes on test scores, teachers' marks and locus of control. Alternative definition of Actual Multigrade2

	(1)	(2)	(3)	(4)	(5)
	<i>Literacy Score</i>	<i>Numeracy Score</i>	<i>Teacher Mark Literacy</i>	<i>Teacher Mark Numeracy</i>	<i>External Locus of Control</i>
Panel A: Two-Stage Least Squares Estimates					
<i>Actual Multigrade2</i>	-4.916*	-11.457**	-0.458*	-0.451*	0.428
	(2.948)	(5.044)	(0.247)	(0.249)	(0.282)
<i>Class Size</i>	-0.203	-0.596	-0.043**	-0.050**	0.032
	(0.260)	(0.430)	(0.020)	(0.020)	(0.025)
Observations	14290	14290	13200	13252	14119
Panel B: First Stage					
Predicted Multigrade	0.233***	0.233***	0.218***	0.222***	0.234***
	(0.014)	(0.014)	(0.015)	(0.015)	(0.014)
Class Size	-0.040***	-0.040***	-0.041***	-0.040***	-0.040***
	(0.003)	(0.003)	(0.003)	(0.003)	(0.003)
First Stage F Statistics	30.927	30.927	24.673	26.027	31.165

Notes: The Table reports TSLS estimates. Standard errors (reported in parentheses) are corrected for heteroscedasticity. In all the regressions we control for our full set of controls and for regional dummies (20 categories). The symbols ***, **, * indicate that coefficients are statistically significant, respectively, at the 1, 5, and 10 percent level.

²³ It is, however, worth noticing that also when using ActualMultigrade2 some 5th graders in single grade classrooms might end up in the treatment group, and then if the effect of being in a multigrade classroom is negative, also such specification underestimates the negative effect.

Finally, instead of using the sample selection procedure described in Section 2, we focused only on schools that have no more than a classroom for the 5th grade and on municipalities with no more than one school and with less than 4000 inhabitants (35th percentile of the population distribution in our sample). In addition, instead of including population size among control variable, as in previous estimates, we use it to instrument class size. Even in this restricted sample we find results (not reported) that are consistent with those discussed above.

8. Concluding remarks

Multigrade classes allow schools to remain located closer to the families they serve and provide their services at a reasonable cost. However, cost-saving considerations are to be evaluated together with the effects that multigrade classes may produce on student outcomes. In this paper we provide additional evidence of the effects of multigrade classes both on cognitive and non-cognitive skills. We exploit the discontinuous rules that regulate class composition in Italy as a source of exogenous variation in the probability of attending a multigrade class.

We find a negative effect on students' performance both in literacy and numeracy standardized test scores. The effect is particularly pronounced for numeracy test scores: students placed in multigrade classes obtain a lower score (by half standard deviation) when compared to students in single grade classes.

On the contrary, a negative but not statistically insignificant effect is found when looking at marks assigned by teachers. This might depend on the fact that teachers of multigrade classes, aware of the more complex environment faced by their pupils, tend to reward their effort more generously. This is especially true for teachers working in the Southern part of the country, while in schools located in the North multigrade classes harm students' skills also when measured by teacher assessment.

Finally, we show that placing students in multigrade classrooms causes an increase in their tendency to attribute successful and failure situations to factors behind their own control.

Since our results pertain to students attending the 5th grade and who are likely to have attended multigrade classes also in previous years of their educational career (although we do not possess this piece of information), the estimated effects can be interpreted as the cumulative effects of having attended a substantial part of the primary school cycle in a multigrade environment.

This evidence suggests that the use of multigrade classes might produce a number of unintended consequences that are to be managed by school administrators. In order to define effective interventions it would be necessary to better understand the mechanisms that drive these negative effects. They could be the result of excessive teachers' workload or lack of training on teacher practices within this peculiar classroom environment. It could also be that relationships among peers work in an asymmetric way, as suggested by Leveun et al. (2014): in such a case the negative effect we find could be due to having spent the final year with lower grade peers (which on the contrary could have benefited by being exposed to older peers).

Understanding which of these channels slows down the learning process of students placed in multigrade classes is an interesting topic for future research.

Appendix

In Table A.1 are reported descriptive statistics for the entire population of 5th grade students 2011-2012.

Table A.1. Descriptive statistics. Whole sample

Panel (a). Whole sample					
	Mean	St. Dev.	Min	Max	Obs.
<i>Actual Multigrade1</i>	0.010	0.100	0	1	466,897
<i>Actual Multigrade2</i>	0.019	0.136	0	1	466,897
<i>Predicted Multigrade</i>	0.020	0.139	0	1	473,857
<i>Rasch Literacy Score</i>	202.669	40.588	-35.360	368.097	450,179
<i>Literacy Score</i>	74.448	18.587	0	100	473,725
<i>Teacher markLiteracy</i>	7.623	1.148	1	10	408,339
<i>Rasch Numeracy Score</i>	186.196	42.464	2.281	367.182	445,513
<i>Numeracy Score</i>	52.710	21.229	0	100	473,749
<i>Teacher mark Numeracy</i>	7.701	1.189	1	10	408,185
<i>External Locus of Control</i>	2.202	1.641	0	8	457,148
<i>External Locus of Control Positive</i>	1.310	1.138	0	5	461,126
<i>External Locus of Control Negative</i>	0.895	0.951	0	3	465,192
<i>Female</i>	0.498	0.500	0	1	473,857
<i>Regularly Enrolled</i>	0.949	0.219	0	1	473,579
<i>Early Enrolled</i>	0.017	0.131	0	1	473,579
<i>Late Enrolled</i>	0.033	0.179	0	1	473,579
<i>Pre Primary School</i>	0.182	0.386	0	1	473,857
<i>ESCS index</i>	0.142	1.026	-3.081	2.722	452,782
<i>Italian Father</i>	0.805	0.396	0	1	473,857
<i>Italian Mother</i>	0.797	0.402	0	1	473,857
<i>Full day</i>	0.266	0.442	0	1	453,341
<i>Number of student enrolled in grade 5th</i>	20.272	4.318	1	34	473,857
<i>Class size</i>	20.284	4.281	1	34	473,857
<i>Southern regions</i>	0.384	0.486	0	1	473,857
<i>Altitude</i>	172.519	194.45	0	2035	471,117
<i>Population</i>	210323.8	550179	106	2617175	471,117

Notes: The data are drawn from the Invalsi websites (downloaded on 7/2/2016).

In Table A.2, Panel (a), we report descriptive statistics for the sample of 5th grade students attending schools that have no more than a classroom in the 5th grade. Panel (b) of Table A.2 split this population into students placed in single grade classes (Panel (b)) and students in a multigrade class (Actual Multigrade1 – panel (c)).

Table A.2. Descriptive statistics

	Panel (a) Whole sample			Panel (b) Single Grade			Panel (c) Actual Multigrade1		
	Mean	St. Dev.	Obs.	Mean	St. Dev.	Obs.	Mean	St. Dev.	Obs.
<i>Actual Multigrade1</i>	0.042	0.202	105,442						
<i>Actual Multigrade2</i>	0.078	0.269	105,442						
<i>Predicted Multigrade</i>	0.058	0.233	105,442	0.030	0.169	100,963	0.695	0.461	4,479
<i>Rasch Literacy Score</i>	202.246	41.058	100,003	202.577	41.039	95,953	194.399	40.729	4,050
<i>Literacy Score</i>	74.229	18.649	105,442	74.451	18.469	100,963	69.210	21.732	4,479
<i>Teacher Mark Literacy</i>	7.583	1.123	95,998	7.587	1.124	91,805	7.495	1.085	4,193
<i>Rasch Numeracy Score</i>	182.933	42.393	98,847	183.568	42.263	94,813	168.021	42.708	4,034
<i>Numeracy Score</i>	52.541	21.278	105,442	52.696	21.257	100,963	49.041	21.441	4,479
<i>Teacher Mark Numeracy</i>	7.672	1.171	96,090	7.676	1.173	91,880	7.584	1.124	4,210
<i>External Locus of Control</i>	2.213	1.640	102,128	2.210	1.640	97,790	2.291	1.654	4,338
<i>Ext Locus Control Positive</i>	1.326	1.145	102,927	1.324	1.144	98,558	1.371	1.150	4,369
<i>Ext.Locus Control Negative</i>	0.891	0.948	103,759	0.890	0.947	99,351	0.924	0.973	4,408
<i>Female</i>	0.498	0.500	105,442	0.499	0.500	100,963	0.488	0.500	4,479
<i>Regularly Enrolled</i>	0.950	0.218	105,442	0.950	0.218	100,963	0.953	0.211	4,479
<i>Early Enrolled</i>	0.017	0.130	105,442	0.018	0.131	100,963	0.010	0.101	4,479
<i>Late Enrolled</i>	0.033	0.178	105,442	0.033	0.178	100,963	0.036	0.187	4,479
<i>Pre Primary School</i>	0.162	0.369	105,442	0.165	0.372	100,963	0.090	0.286	4,479
<i>ESCS index</i>	0.152	1.014	105,442	0.162	1.016	100,963	-0.053	0.927	4,479
<i>Italian Father</i>	0.853	0.354	105,442	0.853	0.354	100,963	0.860	0.347	4,479
<i>Italian Mother</i>	0.841	0.366	105,442	0.841	0.366	100,963	0.835	0.371	4,479
<i>Full day</i>	0.207	0.405	105,442	0.209	0.407	100,963	0.156	0.363	4,479
<i># student enrolled in grade 5</i>	18.907	5.466	105,442	19.392	5.013	100,963	7.954	3.389	4,479
<i>Class size</i>	18.956	5.357	105,442	19.393	5.010	100,963	9.087	2.834	4,479
<i>Southern regions</i>	0.316	0.465	105,442	0.314	0.464	100,963	0.366	0.481	4,479
<i>Altitude</i>	234.368	244.022	105,442	222.480	234.360	100,963	502.334	297.930	4,479
<i>Population</i>	132499.2	440070.1	105,442	137955.2	448614.6	100,963	9513.375	81882.18	4,479

Notes: The data are drawn from the Invalsi websites (downloaded on 7/2/2016).

In Table A.3 we report descriptive statistics for the sample used in our main analysis considering a different definition of being in a multigrade class (Actual Multigrade2).

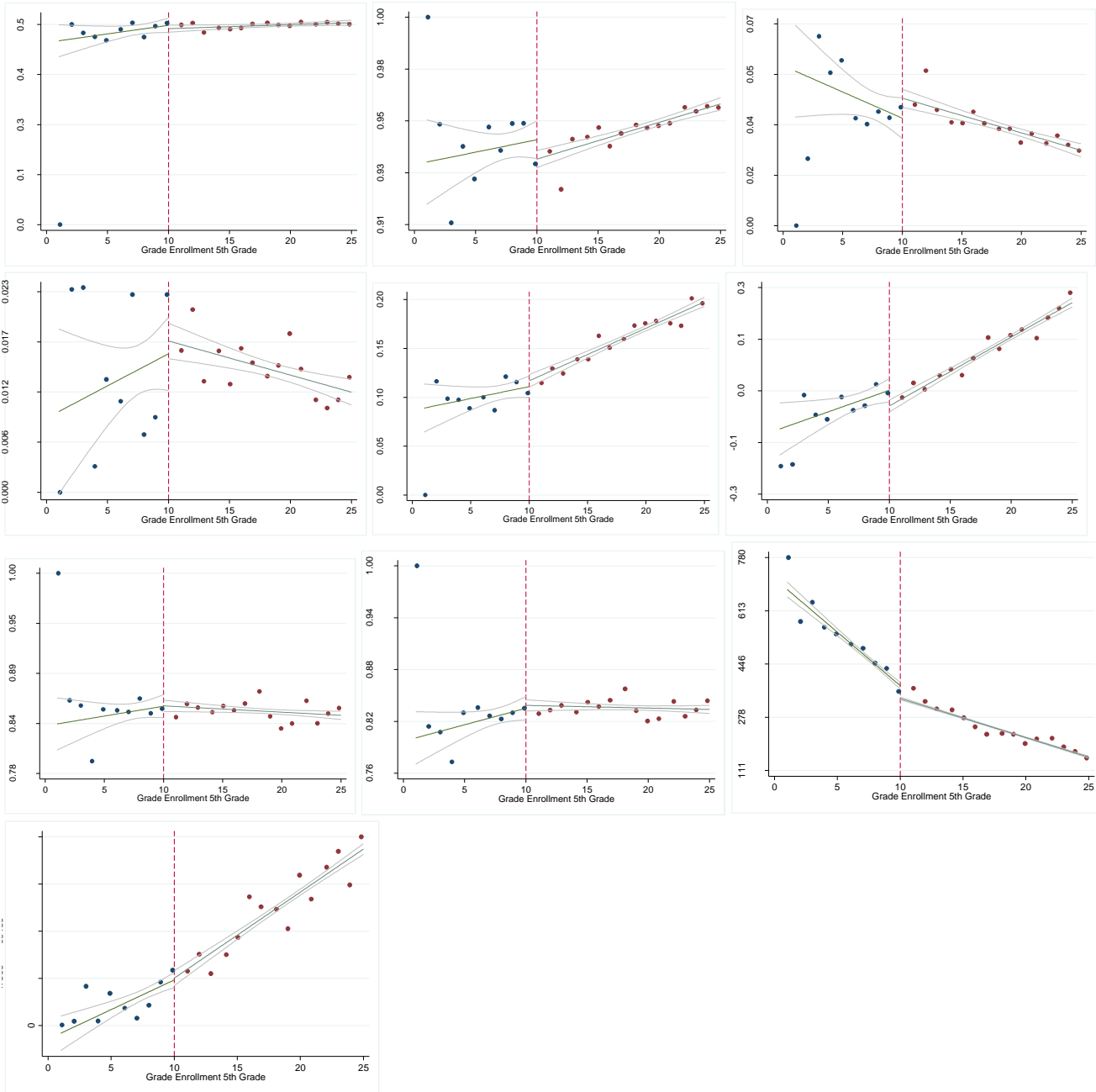
Table A.3. Descriptive statistics – Sample used in the main analysis

	Panel (b) Single Grade			Panel (c) Actual Multigrade2		
	Mean	St. Dev.	Obs.	Mean	St. Dev.	Obs.
<i>Predicted Multigrade</i>	0.167	0.373	8,756	0.575	0.494	5,534
<i>Rasch Literacy Score</i>	197.802	41.242	8,756	196.311	41.950	5,534
<i>Literacy Score</i>	75.826	13.862	8,756	75.252	14.071	5,534
<i>Teacher Mark Literacy</i>	7.468	1.111	8,057	7.516	1.105	5,143
<i>Rasch Numeracy Score</i>	175.082	41.018	8,756	171.604	41.968	5,534
<i>Numeracy Score</i>	53.378	19.489	8,756	52.254	19.767	5,534
<i>Teacher Mark Numeracy</i>	7.557	1.152	8,076	7.609	1.152	5,176
<i>External Locus of Control</i>	2.212	1.669	8,648	2.301	1.662	5,471
<i>Ext Locus Control Positive</i>	1.347	1.162	8,648	1.390	1.162	5,379
<i>Ext.Locus Control Negative</i>	0.893	0.961	8,516	0.942	0.968	5,426
<i>Female</i>	0.489	0.500	8,580	0.495	0.500	5,534
<i>Regularly Enrolled</i>	0.936	0.244	8,756	0.952	0.214	5,534
<i>Early Enrolled</i>	0.018	0.134	8,756	0.010	0.099	5,534
<i>Late Enrolled</i>	0.045	0.208	8,756	0.038	0.191	5,534
<i>Pre Primary School</i>	0.130	0.337	8,756	0.091	0.287	5,534
<i>ESCS index</i>	0.001	0.994	8,756	- 0.024	0.937	5,534
<i>Italian Father</i>	0.849	0.358	8,756	0.852	0.355	5,534
<i>Italian Mother</i>	0.833	0.373	8,756	0.829	0.377	5,534
<i>Full day</i>	0.169	0.375	8,756	0.148	0.355	5,534
<i># student enrolled in grade 5</i>	11.265	1.800	8,756	8.718	2.768	5,534
<i>Class size</i>	11.265	1.800	8,756	9.431	2.185	5,534
<i>Southern regions</i>	0.405	0.491	8,756	0.341	0.474	5,534
<i>Altitude</i>	327.294	278.921	8,756	468.7481	295.627	5,534
<i>Population</i>	703.289	316.9525	8,756	74,210.49	34.849	5,534

Notes: The data are drawn from the Invalsi website (downloaded on 7/2/2016)

In the following Figures we present a number of school and student characteristics plotted against the number of students enrolled in 5th grade. Each figure depict enrollment in 5th grade cell means of the predetermined characteristics in the proximity of minimum class size, along with the fitted values of a locally weighted regression which is calculated within each segment. In general, the figures only show small differences at the threshold point.

Figure A1. Discontinuity in the predetermined characteristics



In Table A.4 we report the OLS estimates corresponding to the IV model estimated in Tables 3 and 4.

Table A.4 OLS Estimates of Multigrade Classes on Literacy and Numeracy Scores

	(1)	(2)	(3)	(4)
	<i>Literacy Score</i>	<i>Rasch Literacy Score</i>	<i>Numeracy Score</i>	<i>Rasch Numeracy Score</i>
<i>Actual Multigrade</i>	-1.207*** (0.437)	-2.274* (1.349)	-1.076 (0.746)	-5.483*** (1.442)
<i>Class Size</i>	0.102 (0.097)	0.506* (0.283)	0.252* (0.144)	0.780*** (0.294)
<i>Female</i>	-2.674*** (0.231)	-7.813*** (0.696)	2.216*** (0.330)	2.707*** (0.581)
<i>Pre Primary School</i>	-0.144 (0.397)	-0.308 (1.171)	0.240 (0.588)	3.243** (1.294)
<i>Early Enrolled</i>	-1.759* (1.049)	-4.715 (3.127)	-0.116 (1.566)	7.139** (3.210)
<i>Late Enrolled</i>	-6.638*** (0.744)	-17.345*** (1.978)	-6.564*** (0.874)	-6.988*** (1.642)
<i>Full day</i>	-0.389 (0.567)	-0.792 (1.674)	0.829 (0.812)	-1.173 (1.705)
<i>ESCS index</i>	2.858*** (0.142)	9.488*** (0.417)	4.127*** (0.199)	5.140*** (0.420)
<i>Italian Father</i>	2.145*** (0.445)	6.356*** (1.304)	2.004*** (0.648)	2.682** (1.192)
<i>Italian Mother</i>	1.321*** (0.398)	5.223*** (1.190)	1.566*** (0.583)	1.355 (1.102)
<i>Altitude</i>	0.001 (0.001)	0.005* (0.003)	0.004*** (0.001)	-0.003 (0.003)
<i>Population</i>	-0.000 (0.001)	0.001 (0.003)	-0.000 (0.002)	0.011*** (0.003)
Observations	14290	14290	14290	14290

Notes: The Table reports OLS estimates. Standard errors (reported in parentheses) are corrected for heteroscedasticity. In all the regressions we control for regional dummies (20 categories, not reported). The symbols ***, **, * indicate that coefficients are statistically significant, respectively, at the 1, 5, and 10 percent level.

The effects of multigrade classes on Literacy and Numeracy Scores: estimates considering the whole sample

In this section we run our regressions considering the whole sample of 5th grade students attending schools that have no more than one classroom in the 5th grade. This choice has been made in order to reduce the measurement error in our indicator of students who are effectively attending a multigrade class (Actual Multigrade 1) and in order to limit endogeneity problems in class size.

In Table A.5 we reported the TSLS estimates of the impact of being placed in a multigrade class both on Literacy and Numeracy standardized test scores. We used the same specifications adopted in Table 3 and Table 4 of the paper. However, now we are able to control not only for class size but also for the number of students enrolled in the 5th grade. Even if this is our preferred specification, results do not change when we only control for class size (results not reported and available upon request).

Table A.5. TSLS Estimates of Multigrade Classes on Literacy and Numeracy Scores. Whole Sample

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
	<i>Literacy Score</i>	<i>Literacy Score</i>	<i>Rasch Literacy Score</i>	<i>Rasch Literacy Score</i>	<i>Numeracy Score</i>	<i>Numeracy Score</i>	<i>Rasch Numeracy Score</i>	<i>Rasch Numeracy Score</i>
Panel A: Two-Stage Least Squares Estimates								
<i>Actual</i>	-12.078***	-8.328***	-11.091***	-6.037**	-10.245***	-8.035***	-23.947***	-5.951*
<i>Multigrade1</i>	(1.451)	(2.365)	(1.802)	(2.896)	(1.319)	(2.183)	(2.240)	(3.645)
<i>Class Size</i>		-0.489 (0.389)		-1.173** (0.483)		-0.133 (0.338)		-0.178 (0.503)
Observations	105442	105442	100003	100003	105442	105442	98847	98847
Panel B: First Stage								
<i>Predicted</i>	0.470***	0.359***	0.476***	0.366***	0.471***	0.359***	0.479***	0.370***
<i>Multigrade</i>	(0.002)	(0.003)	(0.002)	(0.003)	(0.002)	(0.003)	(0.002)	(0.003)
<i>Class Size</i>		0.097*** (0.000)		0.097*** (0.000)		0.097*** (0.000)		0.092*** (0.000)
<i>Students Enrolled 5th Grade</i>		-0.100*** (0.000)		-0.101*** (0.000)		-0.101*** (0.000)		-0.097*** (0.000)
First Stage F Stat.	846.009	407.326	763.467	375.52	846.009	407.326	755.68	373.238

Notes: The Table reports TSLS estimates. Standard errors (reported in parentheses) are corrected for heteroscedasticity. In all the regressions we control for our full set of controls and for regional dummies (20 categories, not reported). The symbols ***, **, * indicate that coefficients are statistically significant, respectively, at the 1, 5, and 10 percent level.

Table A.6. reports the TSLS estimates of the impact of being placed in a multigrade class both on student achievement in Literacy and Numeracy as measured by marks assigned by teachers. We estimate the same specifications presented in Table 5 of the paper (we also control for the number of students enrolled in the 5th grade; even if this is our preferred specification, results do not change when we only control for class size).

Table A.6 TSLS Estimates of Multigrade Classes on grades assigned by teachers

	(1)	(2)	(3)	(4)
	<i>Teacher Mark Literacy</i>	<i>Teacher Mark Literacy</i>	<i>Teacher Mark Numeracy</i>	<i>Teacher Mark Numeracy</i>
Panel A: Two-Stage Least Squares Estimates				
<i>Actual Multigrade1</i>	0.005 (0.046)	0.194** (0.077)	0.043 (0.048)	0.268*** (0.079)
<i>Class Size</i>		-0.009 (0.012)		-0.016 (0.012)
Observations	95998	95998	96090	96090
Panel B: First Stage				
<i>Predicted Multigrade</i>	0.473*** (0.003)	0.360*** (0.003)	0.473*** (0.003)	0.360*** (0.003)
<i>Class Size</i>		0.096*** (0.001)		0.096*** (0.001)
<i>Students Enrolled 5th Grade</i>		-0.100*** (0.001)		-0.100*** (0.001)
First Stage F Stat.	1206.12	596.79	1208.22	595.166

Notes: The Table reports TSLS estimates. Standard errors (reported in parentheses) are corrected for heteroscedasticity. In all the regressions we control for our full set of controls and for regional dummies (20 categories, not reported). The symbols ***, **, * indicate that coefficients are statistically significant, respectively, at the 1, 5, and 10 percent level.

In Table A.7 we report the TSLS estimates of the impact of being placed in a multigrade class on our measures of external locus of control. We replicate the same specification introduced in Table 6 of the main text. However, now we are able to control not only for class size but also for the exact number of students enrolled in the 5th grade. Even if this is our preferred specification, results do not change when we only control for class size (results not reported and available upon request).

Table A.7. TSLS Estimates of Multigrade Classes on External Locus of Control. Whole sample

	(1)	(2)	(3)	(4)	(5)	(6)
	<i>External</i>	<i>External</i>	<i>External Neg</i>	<i>External Neg</i>	<i>External Pos</i>	<i>External Pos</i>
Panel A: Two-Stage Least Squares Estimates						
<i>Actual Multigrade2</i>	0.106* (0.058)	0.021 (0.092)	0.086** (0.033)	0.066 (0.052)	0.017 (0.038)	-0.054 (0.060)
<i>Class Size</i>		-0.013 (0.016)		0.010 (0.009)		-0.024** (0.011)
Observations	102,128	102,128	103,759	103,759	102,927	102,927
Panel B: First Stage						
Predicted Multigrade	0.472*** (0.002)	0.360*** (0.003)	0.472*** (0.002)	0.360*** (0.003)	0.472*** (0.002)	0.360*** (0.003)
Class Size		0.097*** (0.000)		0.097*** (0.000)		0.097*** (0.000)
Students Enrolled 5 th Grade		-0.100*** (0.001)		-0.100*** (0.001)		-0.100*** (0.001)
First Stage F Stat-	840.461	402.800	844.850	405.611	841.978	404.108

Notes: The Table reports TSLS estimates. Standard errors (reported in parentheses) are corrected for heteroscedasticity. In all the regressions we control for our full set of controls and for regional dummies (20 categories, not reported). The symbols ***, **, * indicate that coefficients are statistically significant, respectively, at the 1, 5, and 10 percent level.

The effects of multigrade classes on Literacy and Numeracy Scores: estimates considering 2nd Grade students

In Tables A8 and A9 we report estimates obtained when co. We run the same specifications reported in Tables 3 and 4 in the main test of the paper. Instead of controlling for the *ESCS* index (not available for students attending the 2nd grade) we control for father and mother education and for parents employment condition.

Table A8. TSLS Estimates of Multigrade Classes on Literacy Scores. 2nd Grade students

	(1)	(2)	(3)	(4)
	<i>Literacy Score</i>	<i>Literacy Score</i>	<i>Rasch Literacy Score</i>	<i>Rasch Literacy Score</i>
Panel A: Two-Stage Least Squares Estimates				
<i>Actual Multigrade1</i>	-2.229* (1.352)	-3.033 (1.949)	-11.924*** (3.068)	-9.835** (4.567)
<i>Class Size</i>		-0.089 (0.179)		0.232 (0.410)
Observations	11,272	11,272	11,272	11,272
Panel B: First Stage				
Predicted Multigrade	0.391*** (0.007)	0.427*** (0.037)	0.391*** (0.007)	0.427*** (0.037)
Class Size		-0.013*** (0.008)		-0.013*** (0.008)
First Stage F Statistics	241.392	128.884	241.392	128.884

Notes: The Table reports TSLS estimates. Standard errors in parentheses are corrected for heteroscedasticity. In all regressions we control for regional dummies (20 categories, not reported). The symbols ***, **, * indicate that coefficients are statistically significant, respectively, at the 1, 5, and 10 percent level.

Table A9. TSLS Estimates of Multigrade Classes on Numeracy Scores. 2nd Grade students

	(1)	(2)	(3)	(4)
	<i>Numeracy Score</i>	<i>Numeracy Score</i>	<i>Rasch Numeracy Score</i>	<i>Rasch Numeracy Score</i>
Panel A: Two-Stage Least Squares Estimates				
<i>Actual Multigrade1</i>	-0.783	-0.311	-10.704***	-10.290**
	(1.890)	(2.834)	(3.562)	(5.224)
<i>Class Size</i>				
Observations	11272	11272	11272	11272
Panel B: First Stage				
<i>Predicted Multigrade</i>	0.391***	0.427***	0.391***	0.427***
	(0.007)	(0.037)	(0.007)	(0.037)
<i>Class Size</i>		-0.013***		-0.013***
		(0.008)		(0.008)
First Stage F Statistics	241.392	128.884	241.392	128.884

Notes: The Table reports TSLS estimates. Standard errors in parentheses are corrected for heteroscedasticity. In all regressions we control for regional dummies (20 categories, not reported). The symbols ***, **, * indicate that coefficients are statistically significant, respectively, at the 1, 5, and 10 percent level.

References

- Almlund, M., A. Duckworth, J. Heckman and T. Kautz (2011), Personality psychology and economics. *Handbook of the Economics of Education*, 4, pp. 1-181
- Baron J. and Cobb-Clark D., (2010), Are young people's educational outcomes linked to their sense of control?
- Bertoni M., G. Brunello, 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, pp. 65-77
- Caliendo M. D. Cobb-Clark and A. Uhlendorff (2015), Locus of Control and Job Search Strategies, *Review of Economics and Statistics*, 97(1), pp. 88-103
- Cebi M. (2007), Locus of Control and Human Capital Investment Revisited, *Journal of Human Resources*, 4, pp. 919-932.
- Chiteji N. (2010), Time-preference, non-cognitive skills and well-being across the life course: Do non-cognitive skills encourage healthy behavior? *American Economic Review*, 100(2), pp. 200–204
- Coleman M. and DeLeire T. (2003), An Economic Model of Locus of Control and the Human Capital Investment Decision, *Journal of Human Resources*, 38(3), pp. 701-721.
- Dardanoni, V., S. Modica and A. Pennisi (2009), Grading across schools. *The B.E. Journal of Economic Analysis & Policy*, 9(1), Article 16.
- Gatz M. and M. Karel (1993), Individual Change in Perceived Control over 20 Years, *International Journal of Behavioral Development* 16(2), pp.305-322
- Gerhardts, I., U. Sundeand L. Zierow (2016), Denominational Schools and Returns to Education, Gender Socialization in Multigrade Classrooms? Beiträge zur Jahrestagung des Vereins für Socialpolitik 2016: Demographischer Wandel - Session: Schooling and Labor Supply, No. A14-V1
- Groves M. (2005), How important is your personality? Labor market returns to personality for women in the US and UK, *Journal of Economic Psychology*, 26(6), pp. 827-841
- Hoxby, C.M. (2000). 'Peer effects in the classroom: learning from gender and race variation', Working Paper 7867, NBER.
- Lindström E. and E. Lindahl (2011). The effects of mixed-age classes in Sweden, *Scandinavian Journal of Educational Research*, 55(2), pp.121-144.
- Leuven, E and M. Rønning (2014). Classroom Grade Composition and Pupil Achievement. *Economic Journal*, 126(593), pp.1164-1192
- Mason, D.A. and R. Burns (1997). 'Reassessing the effects of combination classes', *Educational Research and Evaluation*, vol. 3(1), pp. 1–53.
- Mulkeen A. and C. Higgins (2009) Multigrade Teaching in Sub-Saharan Africa: Lessons from Uganda, Senegal, and the Gambia, World Bank Publications.
- Paccagnella M. and Sestito P., 2014, School cheating and social capital, *Education Economics* 22, pp. 367-388.
- Ricci R. (2010). The Economic, Social, and Cultural Background: a continuous index for the Italian Students of the fifth grade, mimeo
- Sattari R. (2016). The Effect of Multigrade Classrooms on Student Behavior, mimeo.
- Schurer, S. (2011), Does attitude matter? Health, labor supply, and locus of control. mimeo.
- Thomas, J. L. (2012), Combination classes and educational achievement, *Economics of Education Review* 31(6), pp.1058-1066.
- Veenman S. (1995). Cognitive and Noncognitive Effects of Multigrade and Multi-Age Classes: A Best-Evidence Synthesis, *Review of Educational Research*, 65(4), pp. 319-381