

Relative Age, Class Assignment and Academic Performance: Evidence from Brazilian Primary Schools*

Martin Foureaux Koppensteiner
University of Leicester, Leicester, LE1 7RH, UK
mk332@le.ac.uk

Abstract

Students in Brazil are typically assigned to classes based on the age ranking in their cohort. I exploit this rule to estimate the effects on maths achievement of being in class with older peers for students in fifth grade. I find that being assigned to the older class leads to a drop in Math scores of about 0.4 of a standard deviation for students at the cut-off. I provide evidence that heterogeneity in age is an important factor behind this effect. Information on teaching practices and student behaviour sheds light on how class heterogeneity harms learning.

JEL classification: I20, I21.

Keywords: Primary education, group effects, group heterogeneity, regression discontinuity.

* I am very grateful to Francesca Cornaglia, Claudio Ferraz, Randi Hjalmarsson, Marco Manacorda, Barbara Petrangelo, Rodrigo Soares, and seminar participants at PUC Rio, Queen Mary, Centre for Economic Performance LSE, Alicante, Leicester, ZEW Mannheim, the Royal Economic Society Meeting, ESPE, the North American Winter Meeting of the Econometric Society, the EALE/SOLE 3rd International Conference, the IZA Summer School in Labor Economics, and the Congress of the European Economic Association for very useful comments. I am also very grateful to two anonymous referees for suggestions that have substantially improved this manuscript. This is a substantially revised version of a paper previously circulated with the title “Class Assignment and Peer Group Effects: Evidence from Brazilian Primary Schools”. I would like to thank the Secretariat of Education in Minas Gerais, the Brazilian Ministry of Education and the National Institute for Educational Studies and Research (INEP) for providing me with the data. The usual disclaimer applies.

I. Introduction

The question of whether a group composition matters for the outcome of an individual member of that group has received considerable attention in numerous contexts where social interactions may be present. Peer effects have been studied in the context of schools, universities, workplaces, neighbourhoods and prisons among other institutions.¹ Due to the natural grouping of students into schools and classrooms, and the potential for education policies to affect the peer group composition, peer effects in education have received extensive attention from economists. Recent work goes beyond linear-in-means specifications and points to the potential relevance of the distribution of peer characteristics in explaining group effects (Hoxby and Weingarth 2006, Lyle 2009).

The identification of group effects is challenging, due to conceptual problems as well as data limitations. In the education sphere, for example, an identification strategy for peer effects needs to address a potential endogenous selection of students into schools and classes. With selection into groups, unobserved characteristics such as ability, parental support and students' effort are likely to be correlated among peers, and educational outcomes are therefore correlated within the peer group even in the absence of externalities. In addition, the analysis needs to deal with separating peer effects from common shocks to the peer group, such as differential educational and teacher inputs, and it needs to account for the simultaneous determination of student and peer achievement (Manski 1993, Hanushek et al. 2003).

¹ Recent studies include Mas and Moretti (2009) on productivity effects for supermarket cashiers; Bandiera, Barankay and Rasul (2010) on social networks and worker productivity in farm production; Bayer, Hjalmarsson and Pozen (2009) on the effect of juvenile offenders serving time on others' subsequent criminal behaviour, to name just a few. Studies on peer effects in education include Hoxby (2000) for gender and race peer effects; Hanushek et al. (2003) provide a framework for estimating peer effects trying to overcome omitted variables and simultaneous equation biases; Duflo, Dupas and Kremer (2010) provide evidence from a randomised experiment in Kenya; Lavy, Paserman and Schlosser (2008) look at ability peer effects and potential channels; Lavy, Silva and Weinhardt (2009) study the distributional effects of ability peer effects; Lavy and Schlosser (2011) examine gender peer effects and their operational channels; Zimmerman (2003) and Sacerdote (2003) look at peer effects in college education; Angrist and Lang (2004) study peer effects on racial integration and Ammermueller and Pischke (2009) do a cross-country comparison of peer effects at primary school level. Student tracking, school choice, busing, admission policies, class formation, repetition policies and residential location decisions are relevant policy issues that can change the peer composition in schools and classrooms (Zimmerman 2003 and Hanushek et. al 2003).

Randomised experiments are the first choice for overcoming the selection problem, and there have been a number of recent applications in this area. (See Duflo, Dupas and Kremer (2011) on ability grouping in primary schools, Whitmore (2005) looks at gender peer effects, and Cascio and Schanzenbach (2016) at peer age composition, both using data from Project STAR.) Empirical strategies that exploit natural experiments, such as conditional random assignment of college roommates by Zimmerman (2003) and Sacerdote (2003), or the idiosyncratic variation in the gender or racial composition of a given cohort over time have also been used (Hoxby, 2000). There is little experimental or quasi-experimental evidence that overcomes the identification problems of peer group effects in primary or secondary education and even less evidence that specifically considers distributional features of peer groups that might affect educational achievement.

This paper provides quasi-experimental evidence on peer effects from exogenous variation in group membership by using an assignment mechanism of students into classes, which provides the basis for a regression discontinuity (RD) design. Brazilian primary school students are typically allocated to classes based on their relative age in the cohort. Using the age rank as a continuous assignment variable, this rule creates a discontinuity in the allocation to a class (peer group) for students close to the class size cap of the relatively younger class. I exploit this rule to compare outcomes of students at the margin of being assigned to an older group versus a younger group in schools with two classes per cohort. Because of this allocation mechanism these groups differ widely in terms of average student characteristics.

Using two-stage-least squares to estimate the discontinuity in a fuzzy RD setting, I find strong evidence for sizeable group effects. I estimate a negative effect from being in the relatively older class on maths test scores among students in fifth grade of around half of a standard deviation.

The RD strategy in this setting is non-standard as the cut-off point is school specific so that the discontinuity based on the size of the younger class is potentially endogenous. If students were strategically re-allocated to classes based on their latent outcomes precisely at the discontinuity, the variation in outcomes around the threshold would not be ‘as good as random’

and differences in outcomes between those on the right and on the left of the cut-off would not provide consistent estimates of the parameter of interest (Lee and Lemieux 2010). In the paper, though, I argue that assignment to the groups is largely predetermined (in 1st grade) and I find no evidence, based on a large array of observable covariates, of non-random sorting around the proposed cut-off point.²

Because I have data on more than 350 schools, I am able to estimate a separate parameter for each school and relate the magnitude of the estimated coefficient to differences in class characteristics across schools. This strategy allows me to learn about which observable differences across classes, if any, drive the estimated gap in the attainment between barely eligible and barely ineligible pupils. Because, in Brazil, as in many other low- and middle-income countries, grade repetition is widespread, older classes tend typically to display larger variation in age. I find that differences in the age dispersion between older and younger classes seems to play an important role in explaining the estimated test score gap. I do not find such evidence for differences in other class characteristics, including mean age, mean grades repeated, class size and socio-economic status. The paper also presents evidence on differences in the teaching practices across classes that could be partially induced by the class composition. Students in the older class that are more heterogeneous in age state that their teacher is available less likely to clarify doubts, that the teacher spends more time on some students than others and that they have less opportunity to express their opinion in class. Students in the older class also report more frequently that their peers are noisy and disruptive, and that the teacher needs to wait for noise to settle to start teaching. Heterogeneity of the class composition is one possible explanation for these observed differences in teaching practices and student behaviour. Group heterogeneity has to date not received much attention in the literature on peer effects. It has, though, been addressed in the literature on tracking (also referred to as streaming), where

² Table A2 provides information on the initial assignment of students and the transition from one grade to the next.

students are separated by academic ability into schools or classes.³ Some recent research on the effects of tracking that addresses the endogeneity of tracking decisions finds that tracking may benefit equally students from lower and higher achievement tracks. Figlio and Page (2002) show that tracking may actually help low-ability students without proposing a specific mechanism for this effect, and Zimmer (2003) presents quasi-experimental evidence that a negative direct peer effect for low-achieving students is offset by the positive effects of achievement-targeted instruction. Duflo, Dupas and Kremer (2011) use a quasi-experimental assignment of pupils to classes to study the effect of tracking students on initial achievement among Kenyan primary school students. They find persistent positive effects across the achievement distribution of tracking students in a higher and a lower ability class. They attribute this effect mainly to teacher effort and the choice of target teaching level, given the particular incentives for teachers in Kenyan schools, and the better match of the instruction level due to reduced heterogeneity in ability in the classrooms. Their results are matched by the findings of Zimmer (2003) and Hoxby and Weingarth (2006), who show that students in more homogenous classes benefit from more tailored instruction. De Giorgi, Pellizzari and Woolston (2010) provide evidence on the effect of class heterogeneity on academic achievement and labour market outcomes in a higher education setting. They find that the effect of the peer distribution on student performance is non-linear and appears to be inversely U-shaped with respect to the dispersion of gender and ability in the group. The paper contributes to this emerging literature that explicitly considers group heterogeneity in estimating peer effects.

The remainder of this paper is organised as follows: Section 2 briefly describes the Brazilian educational system and the educational system in the state of Minas Gerais, which is the focus of this study. Section 3 describes the data. Section 4 presents the assignment mechanism of students to classes and introduces the identification strategy. Section 5 presents tests for non-

³ There is an extensive pedagogic literature on age, ability grouping and academic tracking. See Robinson (2008), Adams-Byers, Squiller Whitsell and Moon (2004), and Betts and Shkolnik (1999) for some recent examples. Kremer (1997) provides an economic model of sorting.

random sorting and Section 6 presents the main results and for correlated effects. Section 7 gives an interpretation of the peer group estimates and section 8 concludes.

II. The educational system in Brazil and in Minas Gerais state

Primary schooling in Brazil is compulsory and consists of nine years of schooling. Children who turn six years of age by March 31st of a given year are required to commence primary school in that year. The allocation of students to public schools is based on the area of residence in such a way that parents cannot choose a particular school for their children. There exists a sizeable private sector engagement in the provision of primary schooling but, as private institutions charge substantial fees, access is limited to children from middle- and high-income families.⁴ Public schools, in contrast, are free of charge at all ages.

In the public schools of Minas Gerais, which are the focus of this analysis, ‘normal’ class size is set at 25 students per class.⁵ When enrolment per grade is above 25 pupils, the school administration needs to make a choice on how to assign students to classes before the start of the school year. As, unlike innate ability or behavioural characteristics, the age of students at the point of enrolment in first grade can be easily observed by school administrators, age sorting provides a convenient and widely used way of grouping students utilising observable characteristics at the time of entry into primary school.⁶

Students who progress in the usual way typically remain in their original class throughout primary school, so that, other than because of migration between schools and dropouts, assignment to classes is largely predetermined in first grade and not based on any observable characteristics of students other than age.⁷ Obviously, grade repetition may potentially lead to

⁴ Around 10% of schoolchildren in Minas Gerais attend private schools. Source: Brazilian school census 2007.

⁵ Law 16.056 of 24th April 2006 limits class size to 25 students in the initial years of primary education (1st-5th grade) in all public schools in Minas Gerais. Exceptions are theoretically only allowed under special circumstances and during the transitional period of the introduction of the law (<http://goo.gl/bPtsV7>).

⁶ Grouping students according to their age may in fact at least partially coincide with grouping according to ability, as ability is likely to be correlated with age at time of primary school enrolment. See Cascio and Whitmore Schanzenbach (2016) and Angrist and Krueger (1991) for a discussion of student age and educational outcomes.

⁷ Appendix A2 provides more information on the initial assignment of students.

Table 1: Means and proportions of student and teacher characteristics

Panel A: Class and student characteristics		Younger class		Older class	
Age	Math score	527.226	(95.128)	474.844	(92.525)
	Class rank	0.360	(0.181)	0.743	(0.262)
	Class size	24.738	(5.477)	21.868	(5.762)
	(in years)	10.930	(0.822)	11.670	(1.125)
Sex	Female	0.524	(0.455)	0.458	(0.459)
Race	White	0.306	(0.461)	0.264	(0.440)
	Mixed	0.526	(0.481)	0.517	(0.500)
	Black	0.097	(0.295)	0.143	(0.349)
	East-Asian	0.027	(0.163)	0.034	(0.179)
Repeater	Indigenous	0.044	(0.206)	0.042	(0.200)
	Never repeated	0.797	(0.394)	0.489	(0.500)
	Repeated once	0.142	(0.353)	0.292	(0.392)
	Repeated twice	0.043	(0.199)	0.148	(0.351)
SES	Repeated 3 or more times	0.018	(0.129)	0.070	(0.248)
	Family with Bolsa Família	0.480	(0.473)	0.592	(0.492)
	Household employs domestic worker	0.137	(0.346)	0.113	(0.319)
	Number of books	23.496	(28.180)	19.428	(26.610)
	Number of cars	0.608	(0.782)	0.503	(0.663)
	Number of computers	0.262	(0.445)	0.195	(0.404)
	Number of fridges	0.999	(0.442)	0.958	(0.468)
	Number of freezers	0.302	(0.538)	0.282	(0.527)
	Number of radios	1.342	(0.703)	1.286	(0.697)
	Number of TVs	1.497	(0.673)	1.396	(0.685)
	Number of DVD players	0.849	(0.616)	0.786	(0.640)
	Number of bathrooms	1.246	(0.557)	1.175	(0.505)
	Number of washing machines	0.758	(0.591)	0.752	(0.565)
	Number of tumble dryers	0.168	(0.426)	0.163	(0.389)
Panel B: Teacher characteristics					
Sex	Female	0.983	(0.172)	0.965	(0.234)
Age	(in years)	40.495	(7.401)	40.094	(7.729)
Race	White	0.456	(0.461)	0.477	(0.483)
	Mixed	0.420	(0.456)	0.399	(0.453)
	Black	0.093	(0.266)	0.081	(0.255)
	East-Asian	0.028	(0.151)	0.039	(0.192)
Highest educational degree	Indigenous	0.004	(0.064)	0.004	(0.063)
	Secondary education	0.100	(0.279)	0.118	(0.299)
	Higher education – pedagogic degree	0.210	(0.374)	0.208	(0.371)
	Higher education - regular	0.410	(0.455)	0.389	(0.457)
	Higher education and teaching qualification	0.203	(0.376)	0.174	(0.350)
	Higher education – other	0.076	(0.251)	0.111	(0.296)
	Earnings (in R\$)	771.74	(361.716)	743.60	(378.580)
	Years of experience in education	14.023	(5.599)	13.862	(5.959)
	Participation in continuing education	0.375	(0.438)	0.363	(0.461)

Notes: The data from the upper panel are taken from the student background questionnaires, the data from the lower panel are from the teacher questionnaires. Number of observations: 16,031. Source: PROEB 2007.

changes in the original class assignment. Although grade repetition has been reduced by the introduction of automatic grade promotion in Minas Gerais, Table 1 shows that there still exist a substantial number of students who have repeated at least one school grade. Grade repeaters in first grade are, consistent with an assignment rule based on the age ranking of students in the cohort, usually allocated to the older class when repeating the grade in the following year. In succeeding grades, repeaters regularly are allocated to the older class as well. The propensity for repetition in subsequent grades is, nevertheless, also higher in the older classes, so that the in- and outflow of students into the classes largely cancel out each other and class size is, hence, unaffected by repetition.

III. Data and descriptive statistics

For the purpose of this analysis, I use standardised test scores in mathematics of primary school students in public schools in Minas Gerais, a state in the southeast of Brazil and the second most populous state of the country. Educational standards in Minas Gerais are among the highest of the Brazilian states.⁸ The primary source of data in this study is PROEB (Programme of Evaluation of Basic Education), which provides maths test scores at the pupil level for all students in 5th grade in the state.⁹ I use the data for 2007, as this is the only year that contains detailed information on students' ages.¹⁰ The test is carried out at all public schools in the state and test scores are standardised to a mean of 500, with a standard deviation of 100. Participation is compulsory at school and at individual levels, confirmed by a high student participation rate (93%). Surveyed pupils also answer a detailed socioeconomic questionnaire, which includes information on sex, month and year of birth, racial background and the socioeconomic background of the family.

In the following, I restrict the sample to schools with only two classes. This ensures that enough variation is available to identify sizeable group effects for students around the cut-off

⁸ In the SAEB 2005 nation-wide school evaluation system, the mean maths performance of pupils from Minas Gerais was clearly above the Brazilian average, ranking first among the Brazilian states (<http://goo.gl/bgDQTp>).

⁹ PROEB alternates testing students in either maths or Portuguese, with the 2007 tests focusing on maths.

¹⁰ This is also the reason for choosing PROEB over other Brazilian standardised tests – for example SAEB, in which information on age is also not as detailed.

point, in particular with respect to variation in the distributional features of the class composition.¹¹

The data comprises 16,031 students from 363 public primary schools. Table 1 presents summary statistics for these data split by average age in the two classes. The average age of students on the test day in the younger class is 10.93 years and 21.87 years in the older class, which is about nine months above the ‘normal’ age for this grade.¹² This age-grade mismatch is due to a combination of late enrolment and grade repetition. Figures A1 and A2 depict the distribution of age in the younger and older classrooms revealing a long right tail in the distribution, particularly for the older classes. Students at these schools are overwhelmingly from deprived socioeconomic family backgrounds, and 47% of the families of the students at these schools are recipients of *Bolsa Família*, the Brazilian conditional cash transfer programme for poor and very poor families, compared with around 25% in the total population.¹³

PROEB also includes headmaster and teacher questionnaires. The headmaster questionnaire includes questions on the characteristics of the headmaster, such as age, sex and educational background, and questions on the school’s characteristics and its pedagogic strategy. The teacher questionnaire includes questions on individual characteristics, as well as ones on the students in class.

For part of the analysis on the initial class assignment in the annex, I complement the analysis with data from the 2007 School Census, which was conducted by the National Institute for the Study and Research on Education (INEP) on behalf of the Federal Ministry of Education (MEC) and comprises detailed information on school characteristics for all primary schools in Brazil.

¹¹ The focus on schools with two classes also ensures that school administrators cannot establish *special classes* that do not follow the general assignment mechanism. With more than two classes, the school administration may resort to forming separate classes in which students with specific characteristics are grouped, such as grade repeaters, and are separated from the other students in the cohort, which is not observable to the econometrician. As these *special classes* tend to be rather small, measures of age variation are also more susceptible to outliers (Lyle 2009).

¹² The normal age for students in grade five without late enrolment and repetition should be between 120 and 132 month.

¹³ Families are eligible for *Bolsa Família* if per capita family income is not above R\$120 per month (‘moderately poor’) (US\$63 at 1st June 2007) and receive a monthly R\$20 per child under the condition of regular school attendance and participation in vaccination campaigns. Families below a per capita income of R\$60 (‘extremely poor’) receive an additional basic family allowance of R\$62. See <http://goo.gl/iB1GW> and Lindert et al. (2007) for details.

The data appendix provides detailed information on the data sources and the variables used. Summary statistics from the census for the schools used in this analysis are presented in Table A2 in the online appendix.

IV. Empirical strategy

The identification strategy used in this paper exploits the discontinuity in the assignment rule of students in schools with two classes. The treatment assignment mechanism is based on the value of an observed and continuous variable, the age rank (n) of the individual student in each school, in such a way that the probability of receiving treatment is a discontinuous function of that variable at the class size cap \bar{N}_s , the size of the youngest class.¹⁴

Consider a simple reduced-form model of school achievement

$$Y_{is} = \delta_0 + \delta_1 T_i + f(n) + \varepsilon_i \quad (1)$$

where Y_{is} denotes the outcome variable maths test score for individual i in school s , T_i is the treatment indicator that takes a value of 0 for individuals in the younger class and 1 for individuals in the older class, and ε_i is an individual unobserved error component. I ignore at this stage any covariates one might want to include in the specification to reduce sampling variability in the estimator. Educational achievement measured in terms of test scores is assumed to depend on a smooth function $f(\cdot)$ of the student's age rank, and on being in either the younger or older class indicated by T_i . I employ two-stage least squares to estimate δ_1 , the coefficient of interest, using the discontinuity at the class cap as an instrument for treatment T_i (being in the older class).

In a first stage-equation, I assume that T_i is a function of age rank of students in the school cohort and a dummy D_{is} for being above or below the school-specific discontinuity point \bar{N} given by the maximum class size rule:

¹⁴ Using a 50:50 rule to determine a discontinuity in class membership unfortunately does not provide a sharp enough discontinuity across all schools. Because class size may change after the original allocation in first grade, I allow for a school specific discontinuity point based on the class size of the younger class in 5th grade.

$$T_i = \gamma_1 + \gamma_2 D_{is} + f(n) + \nu_i \quad (2)$$

where ν_i is an error component.

For identification of the class effect δ_1 , a continuity assumption needs to be satisfied, such that student achievement varies continuously with the forcing variable of the age rank in the cohort, outside of its influence through treatment T_i (Lee and Lemieux 2010), so that assignment to either side of the discontinuity threshold is as good as random. In other words, identification of the treatment effect relies on the assumption that just below and above the known cut-off point, individuals are similar in observable and unobservable characteristics, other than being in different classes. In this way, the proposed RD strategy allows me to circumvent the confounding effects induced by non-random sorting of individuals across groups that plagues the literature on spillover effects. For the implementation of the RD strategy, I first rank classes according to average student age and then use the class size of the younger class at fifth grade in each school as the cut-off point for the RD.¹⁵

To gain an understanding on whether schools who allocate students to classes based on their age rank differ systematically from schools who do not, I estimate a linear probability model, where the dependent variable is a binary variable with a value of 1 if student assignment is based on age ranking and zero otherwise and regress this on the rich set of school, headmaster, teacher and students' characteristics.¹⁶ I find little systematic association between the probability of using the age-ranking rule and observable school and pupils characteristics, an exception is size of the school. The results are reported in Table A3 in the online appendix. It seems that with a larger cohort size, administrators are inclined to choose homogenous age sorting whereas the

¹⁵ I use the number of students enrolled in the class at the beginning of fifth grade to determine class size, including additional students that are either repeating the grade or transferring students arriving from other schools, and excluding students that have left the class from the previous grade (either due to grade repetition, drop-out or school transfers).

¹⁶ Specifically, I estimate the following linear model: $Y = \beta_0 + \beta_1 S + \beta_2 D + \beta_3 T + \beta_4 P + u$, where Y takes a value of 1 for an allocation rule that sorts students into homogenous age classes and a value of 0 otherwise. S denotes school characteristics, D headmaster characteristics, T teacher characteristics, P mean characteristics of pupils in the cohort and u an idiosyncratic error term. Table A3 reports the coefficients from the estimated model. Only a few variables are statistically significant at conventional levels: cohort size, the existence of a headmaster's office, the headmaster being of an Asian or indigenous background and the mean number of fridges in student's families.

socioeconomic composition of students and mean teacher characteristics do not seem to be systematically related to the assignment rule of students to classes.

V. Testing for non-random sorting

As already outlined, there are threats to the identification assumption. Although in the present case the forcing variable – age rank – cannot be manipulated the same way as in the setting of a conventional RD design, there are concerns with the potential endogenous setting of the cut-off point. The cut-off used for the RD in this paper is determined by the class size of the younger class and therefore differs across different schools. Although the precise cut-off in terms of the age rank is not likely to be known to parents at time of assignment to classes at first grade, public knowledge of the age-based allocation mechanism and the alleged penalty associated with being assigned to the older class may lead some parents to exert pressure to move their child to the younger class later on. Any such strategic intervention by particularly keen parents only would invalidate the continuity assumption if students precisely above the cut-off were successfully moved to the younger class.¹⁷ If the ability of parents to exert pressure and move their child to the younger class would be systematically related to other unobserved determinants of maths achievement (e.g. the home learning environment or the support the student receives), the assumptions of the RD design may be invalidated.

Similarly, the school administration might manipulate class size in a way to move the youngest student in the older class to the younger class, or vice versa, based on some characteristics that are not necessarily observable to the econometrician and that are correlated with the outcome. In this case, the cut-off point would simply be shifted by one rank upwards or downwards. In reality, this is unlikely to happen, as the allocation of students is decided before classes start at first grade, so that the school administration has no information on the ability,

¹⁷ McCrary (2008) suggests a test for the failure of the random assignment assumption by inspecting for a discontinuity in the density of the forcing variable around the discontinuity point. As the forcing variable in the present case is uniformly distributed due to its nature as a relative rank, this test will not be informative in this analysis.

race or socioeconomic background of the student other than administrative information, such as age or sex, that is to be found in the documents necessary for enrolment, like a birth certificate. Because of the gap between the original assignment to classes in first grade and the SIMAVE test taken in fifth grade, there is also a potential for selective attrition. A bias resulting from selective attrition would likely lead to underestimating the true effect, given that survivors in the older class would need to be better on average compared to survivors in the younger class.

In any of the above instances, if students were selected at the cut-off after assignment to classes, whether by the decision of schools, parental pressure, or attrition, pre-determined characteristics of the students and their families would presumably no longer be balanced on either side of the discontinuity (van der Klaauw 2002).

In the following paragraphs, I use a very rich array of information from the student questionnaire to formally test for the balancing properties of pre-determined student characteristics across the cut-off point. Figure A4 in the online appendix provides a graphical analysis of the balancing properties of baseline covariates by plotting local averages for the covariates, and the local linear regression fits separately on both sides of the threshold. In Figure A4 (part 1), the graphs in columns 1 and 3 plot the individual level probability of being a girl and the probability of self-identifying with different ethnic groups. The fraction of girls reduces smoothly with the age rank. The fraction of white, Asian or indigenous students in the class does not reveal any discontinuity at the threshold, while the fraction of mixed and black students show a minor positive increase at the cut-off point. The average number of months repeated before also does not reveal a discontinuity, but different slopes of the local linear regression fits are apparent, these being induced by the different distribution of repeaters in the two classes. This can be taken as evidence that selective attrition is not a problem in the given context. Columns 1 and 3 of Figure A4 (continued) present the same graphs for a wide range of predetermined socioeconomic characteristics. These variables appear well balanced on both sides of the cut-off point and there is no indication of a discontinuity in the means of these characteristics at the cut-off point. Among two additional proxies for the socioeconomic status of the family, the number

of domestic workers employed and the fraction of families receiving *Bolsa Família*, only the latter shows a small difference around the threshold.

Table 2: RD estimates of individual and family variables

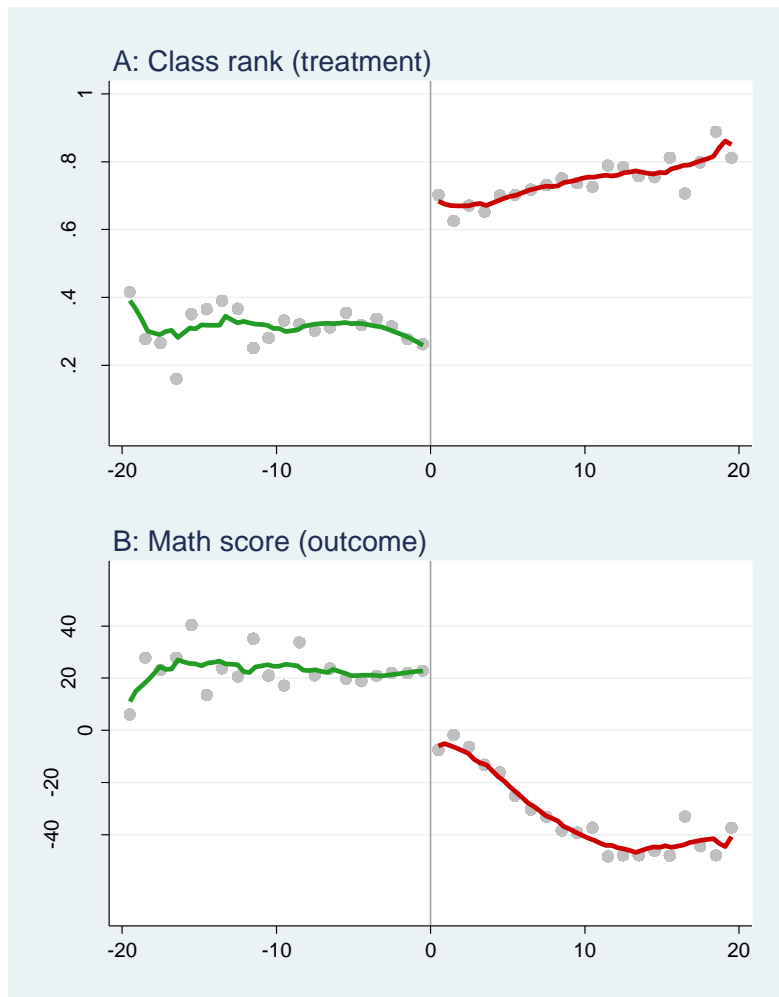
		(1)		(2)	
		Individuals		Peers	
Fraction of:	Age (in months)	0.442	(0.735)	8.157***	(0.796)
	Grades repeated (in months)	0.728	(0.879)	7.487***	(0.457)
	Female	0.190	(0.127)	-0.088***	(0.019)
	White	0.008	(0.092)	-0.035	(0.023)
	Mixed	-0.037	(0.102)	-0.072**	(0.032)
	Black	0.115**	(0.055)	0.089***	(0.018)
	East-Asian	-0.026	(0.022)	0.011	(0.009)
	Indigenous	-0.076	(0.047)	-0.001	(0.009)
	Domestic helper	-0.020	(0.058)	-0.053***	(0.017)
	Bolsa Família	0.165*	(0.099)	0.144***	(0.027)
Number of:	Parental homework support	0.027	(0.054)	-0.066***	(0.016)
	Bathrooms	-0.101	(0.098)	-0.129***	(0.033)
	Books	-4.314	(4.956)	-8.016***	(1.928)
	Cars	-0.167	(0.138)	-0.141***	(0.039)
	Computers	-0.031	(0.068)	-0.108***	(0.022)
	Fridges	0.096	(0.077)	-0.074**	(0.031)
	Freezers	-0.013	(0.087)	-0.052**	(0.025)
	Radios	0.195	(0.158)	-0.083	(0.052)
	Washing machines	0.080	(0.105)	-0.037	(0.033)
	Dryers	-0.057	(0.082)	0.014	(0.021)
	DVDs	0.125	(0.121)	-0.120***	(0.035)
	TV sets	-0.008	(0.141)	-0.194***	(0.042)
	Video players	0.080	(0.107)	-0.066**	(0.028)
Number of student observations		1,688		1,688	

Notes: Entries are separate IV estimates of the class effect on student and family characteristics, where being in the second class has been instrumented by a dummy for having an age rank larger than 0. For each variable a separate regression has been estimated. Column (1) reports the effect around the discontinuity point for the individual values of the characteristics; column (2) reports the estimates for the values of the peer group characteristics for the same individuals around the cut-off point. All specifications include a second-order polynomial in the age rank. Heteroskedasticity consistent standard errors, clustered on the school level are reported in parentheses. *, ** and *** denote significance at the 10%, 5% and 1% level, respectively.

In a formal analysis, I estimate all predetermined characteristics of students using the same specification as for the main estimates in Table 3. Table 2 reports the RD estimates for these variables. Only the estimate for the probability of being a black student is significant, at the 5%

level.¹⁸ None of the other household socioeconomic characteristics reveals a statistically significant difference at the threshold, and most coefficients are small, confirming that the balancing properties of these predetermined characteristics are satisfied. Although the absence of discontinuities in predetermined individual and family characteristics cannot prove the balancing property of unobservables, it is reassuring to find that individuals on both sides of the cut-off are observationally equivalent.

Figure 1: Local averages and local linear regression of treatment and outcome variable



Notes: The graphs plot local averages of the probability of being in older class and of the standardized maths test score according to the age ranking in the cohort as distance of students from the cut-off point and local linear regression fits on both sides of the cut-off point using a rectangular kernel with a bandwidth of 3 months.

¹⁸ Choosing different specifications for the RD by including either only a linear polynomial term or a cubic term makes the estimate for this variable insignificant, so that the single significant estimate can either be attributed to model misspecification or random chance. Any other specification for the functional form or estimating the RD without robust standard errors does not change the significance of the estimates of any of the variables.

In addition, I tested how well predetermined characteristics explain treatment by regressing the treatment indicator on the set of predetermined characteristics. Column 1 of Table A6 in the online appendix reports the coefficients from this regression. Only one of 19 coefficients is significant at the 5 percent level of significance and an F-test rejects the hypothesis for joint significance of these variables.

VI. Results

Before presenting the regression analysis, it is useful to show the raw data. The upper graph of Figure 1 plots the probability of being in the older class in one-month bins, where the age rank has been centred on the cut-off point of zero. The local linear regression fits using a rectangular kernel, with a bandwidth of three months superimposed. The discontinuity in the average class rank at the cut-off point is evident, and the size of the discontinuity in the probability of treatment conditional on the age rank is around 0.5. The estimated increase in the rank is less than one, as not all schools choose to allocate students into homogenous classes.

In panel B of Figure 1, I plot local averages of maths test scores and the local linear regression lines on both sides of the cut-off point. The data show a very clear fall in maths test scores: the oldest pupil in the younger class shows an average attainment in maths that is 0.2 of a standard deviation higher than that of the younger pupil in the older class. Hence, Figure 1 suggests that being assigned to the older class significantly harms learning outcomes.

Table 3 presents the first-stage estimates for the size of the discontinuity in mean class rank, the OLS estimates for the size of the discontinuity in test scores at the discontinuity point and the 2SLS estimates for the causal effect of crossing the cut-off point from the younger class to the older class. All specifications include school-fixed effects that account for observed and unobserved differences across schools that are common across classes. Standard errors are heteroskedasticity consistent and adjusted for clustering at the school level. Column (1) presents the estimates for the models, including only a quadratic polynomial in age rank. Column (2) includes controls for the whole set of predetermined individual and family characteristics. The estimates of column (3) include teacher characteristics in addition to the other covariates.

The top panel of Table 3 presents estimates for the first stage regressions, where the dependent variable is 1 for students being in the older class and zero otherwise. The estimates for the size of the discontinuity range between 0.451 and 0.467, similar to the observed discontinuity in panel A of Figure 1.

Table 3: Main estimation results

	(1)	(2)	(3)
Panel A: first stage			
Dependent variable: class rank			
	0.467***	0.453***	0.451***
	(0.056)	(0.057)	(0.056)
R ²	0.326	0.370	0.403
Panel B: reduced form			
Dependent variable: maths test scores			
	-26.445***	-19.196**	-19.513**
	(7.458)	(7.646)	(7.743)
R ²	0.405	0.482	0.485
Panel C: IV regression discontinuity results			
Dependent variable: maths test scores			
	-56.574***	-42.385***	-43.297***
	(15.299)	(15.455)	(15.673)
R ²	0.410	0.485	0.489
Number of student observations	1,688	1,688	1,688
Window width	1 month	1 month	1 month
Order of polynomial	2	2	2
School fixed effects	yes	yes	yes
Individual controls	no	yes	yes
Teacher controls	no	no	yes

Notes: The top panel reports the first stage regressions using OLS estimating equation (2). The middle panel reports the coefficient on maths test score on the dummy equal 1 for the age rank larger than 0 (reduced form). Test scores are centred using school fixed effects in all specifications. The bottom panel reports IV estimates of the effect of being in the older class on maths test scores, where being in the older class has been instrumented by a dummy for having an age rank larger than 0. All specifications include a second-order polynomial in the age rank and use a window width of 1 month. Specifications in column (2) include the whole set of predetermined individual and family characteristics, including sex, race, repeated years and SES family characteristics; specifications in column (3) additionally include all predetermined teacher characteristics, including teacher sex, race, age, salary, variables on educational background and experience. All estimates use students in one-month bins around the cutoff point. Heteroskedasticity consistent standard errors are clustered by schools and reported in parenthesis. ** and *** denote significance at the 5% and 1% level, respectively.

The middle panel of Table 3 reports the reduced form estimates from an OLS regression, with maths test scores as the dependent variable on a dummy equal to 1 for being to the right of the threshold. Column 1 reports the raw estimate of the discontinuity of maths test scores at the cut-off point.

The bottom panel of Table 3 reports the two-stage-least squares estimates for the class peer effects using the same specifications as for the OLS estimates in panels A and B. The size of the estimated effect, without further controls, is around 0.57 of a standard deviation in maths test scores and significant at the 1% level. Including individual level controls in column 2 reduces the effect by about 25% to around 0.42 of a standard deviation in test scores. The moderate reduction could likely be explained by model misspecification due to the inclusion of the set of predetermined variables (Imbens and Lemieux 2008). The further inclusion of controls for teacher characteristics in column 3 does not affect the estimates notably.¹⁹

Under the identifying assumptions outlined in the previous section, the results can be interpreted as the causal effect on individuals whose treatment status changes, that is, who were to switch from the younger class to the older class as the value of n changes from just below \bar{N} to just above \bar{N} .

Table A1 presents the RD estimates for wider intervals of the discontinuity sample around the cut-off point and different orders of the polynomial terms included in the regressions as a first robustness check. Rows 1 and 2 are the estimates of the RD without any further controls, and rows 3 and 4 are the estimates that have the full set of controls, including individual, family and teacher characteristics. The estimates do not reveal any substantial sensitivity with respect to the choice of the order of the polynomial. Replacing the quadratic by a cubic term leaves the

¹⁹ Formal Hausman tests reject equality of the coefficients for specification (1) and (2) and (1) and (3) at the 5% level of significance. The test does not reject equality of coefficients for specifications (2) and (3) at any conventional level of significance.

estimates virtually unchanged. Increasing the range of observations used for the estimation also does not alter the estimates for the treatment effect in any significant way.

VII. Interpretation of the effects

A crucial question pertains to the channels through which the negative group effect operates. The substantial negative effect could either be driven by direct peer effects, for example, through being with on average lower-performing classmates in the older class, or by indirect effects of the peer group composition that work through behavioural changes by students, teachers or schools to the class composition.

Exogenous peer characteristics

In the literature, it is often assumed that peer characteristics such as sex, race and socioeconomic status are proxies for (unobserved) peer ability and that exogenous peer effects work through being grouped with peers of different ability. The academic achievement of marginal students might be affected because there are more or less bright students who contribute to the learning experience of their peers for example by asking stimulating questions in class.

Column 2 of Table 2 reports the estimates of the difference in mean values of a number of peer variables for students around the cut-off point. The first row reports the difference in peer age in the classrooms and the second row, the difference in mean months repeated by students in the class. Unlike with the individual characteristics, I observe large and significant changes in peers' characteristics at the threshold. Peers in the older class are on average about 8 months older, which is almost completely due to the higher share of repeaters in these classes.²⁰ The remainder is due to late enrolment at first grade and temporary dropout from school followed by re-enrolment later.

Repeaters and students who enrol late at first grade often belong to families from a more deprived socioeconomic background (Patrinos and Psacharopoulos 1996 and Gomes-Neto and Hanushek 1994), which causes the socioeconomic indicators of peers to be systematically

²⁰ Calculation based on the theoretical enrolment age of students and the number of months repeated by students show that repetition accounts for about 75% of total age-grade mismatch.

different between the two classes. The RD estimates for many of these pre-determined characteristics show a statistically significant discontinuity in peer characteristics among students around the cut-off point.

Besides mean age, age dispersion in the class also differs considerably between the two classes. With the larger number of repeaters, age dispersion in the older classes is considerably greater than in the younger classes. The standard deviation of age is about 40% greater (3.6 months) in the older classes (Table 1, row 4). Figures A1 and A2 show the distribution of age of students for the two classes and give a graphical representation of the difference in the distribution of age between the classes.

Overall, students to the right of the cut-off point, while not being different from students just to the left on a range of individual and parental characteristics, have peer groups that not only consist of fewer girls, a higher fraction of blacks, a lower fraction of mixed students and a higher share of children from more deprived socioeconomic background, but also, due to widespread grade repetition, more heterogeneous classmates.

Indirect effects: responses of schools

A concern for the estimation of class peer effects is that correlated effects in the form of common shocks to the peer group (whether exogenous or endogenous) may bias the peer effect estimates. In the present case, one would like to rule out that the negative effect on test scores is not driven by systematically different learning environments provided by the schools to the different classes. Although it is not possible to completely rule out differences in the learning environments across classes as some of these characteristics may be unobservable, I can nonetheless assess whether the observable characteristics, measured by a broad set of teaching resources, teacher and class characteristics, are balanced across classes.

Systematically different learning environments may for example arise from assigning teachers of different quality to either of the two classes. This may happen in a compensatory fashion, such that better teachers are allocated to weaker classes, which would lead to an underestimation of the peer effect. Better educated or more experienced teachers could also be allocated to the

younger class to strengthen good students further, which would lead to overestimating the effect of the peer group. Headmasters are asked in the background questionnaire how they generally allocate teachers to classes. The vast majority (68%) of headmasters report allocating teachers in a non-systematic fashion to classes, either by means of a draw or by no specific criteria. Less than 2% of headmasters

Table 4: Class and teacher characteristics

Dependent variable			
Class characteristics	Std. deviation of age (in months)	4.012***	(0.381)
	Class size	-4.162***	(0.583)
	Non-participation rate (at threshold)	0.006	(0.004)
	Non-participation rate (of peers)	0.093***	(0.022)
Teacher characteristics	Female	-0.087*	(0.049)
	Age (in years)	-1.607	(1.615)
	White	-0.005	(0.101)
	Mixed	-0.048	(0.103)
	Black	0.025	(0.060)
	East-Asian	0.020	(0.033)
	Indigenous	0.009	(0.009)
	Higher education degree	0.030	(0.077)
	Postgraduate degree	-0.034	(0.103)
	Years passed since graduation	-0.108	(0.226)
	Earnings (in Brazilian Reais)	-69.176	(56.943)
	Participation in continuing education	-0.015	(0.091)
	Experience in education (in years)	-0.395	(0.259)
	Teacher has other source of income	-0.089	(0.093)
Teaching resources	Frequency of parent-teacher conferences	0.068	(0.135)
	Quality of textbooks	0.178	(0.098)
	Insufficient financial resources	-0.024	(0.080)
	Insufficient pedagogic resources	-0.063	(0.108)
	Insufficient teaching support staff	0.036	(0.102)
Number of student observations		1,688	

Notes: Entries are separate IV estimates of the class effect on class and teacher characteristics, where being in the second class has been instrumented by a dummy for having an age rank larger than 0. For each variable a separate regression has been estimated. The data come from the teacher questionnaire of PROEB 2007 and the School Census (for class characteristics). Class teacher statements come from the teacher questionnaire and relate to the specific class taught. Class size is calculated using the official number of students enrolled in a class based on information from the School Census. The *non-participation rate (at threshold)* is based on the difference in the distribution of students of age ranks between the school census and PROEB test takers. The *non-participation rate of peers* is based on the difference between class size and number of students participating in the PROEB test. The variable *quality of textbooks* ranges between 0 and 1, with the value 1 given for the best quality and 0 for the lowest. All regressions control for school fixed effects. Heteroskedasticity consistent standard errors are reported in parentheses. * and *** denote significance at the 10% and 1% level, respectively.

allocate more experienced teachers to stronger classes, and around 16% allocate the more experienced teachers to weaker classes. The remainder (13%) allows teachers to select the classes among themselves.²¹ If anything, the teacher allocation would therefore work against finding an effect at the threshold assuming that more experienced teachers would have a positive effect on test scores. To test whether there are indeed any observable systematic differences in teacher characteristics between the younger and older classes, I estimate teacher characteristics for the RD sample of students using the same specification as for the main estimates, and the results are reported in Table 4. None of the teacher characteristics – sex, age, race, experience, education, training and earnings – reveal any significant difference between the two classes, and the estimated coefficients are generally very small, confirming that there are no observable differences in a range of measures of teacher quality across classes. Including teacher characteristics as controls in the RD estimates (Table 3, column 3) also does not change the estimate for the peer effect in any relevant way.²²

Additional information from the teacher questionnaire about the allocation of teaching resources within the school to classes also provides additional evidence that the main estimates are not driven by such common effects. Teachers report on the frequency of parent-teacher conferences, the quality of textbooks and whether the provision of financial and pedagogic resources or of teaching support staff for class teaching is insufficient. None of the variables on teacher characteristics or teaching resources in the classroom, reported in Table 4, is significantly different between the two groups.

As outlined above, there is some concern about the difference in class sizes between the older and younger classes. The estimate in Table 4 reveals that the number of students in the older class is on average lower (by the order of four students) compared to the younger class. As class

²¹ Unlike in settings in which teacher wages are a function of test scores, teacher wages and promotion in public schools in Minas Gerais state are mostly determined by qualification and seniority so that there is less of an economic incentive to teach better classes. Details can be found in law No. 15.293 *Establishing the Careers of Professionals in Basic Education* in the state of Minas Gerais.

²² A formal test does not reject equality of coefficients across specification (2) and (3) in Table 3, where the only difference is the inclusion of teacher controls in specification (3).

size may have an effect on student achievement, this may potentially lead to a bias in the estimation of the peer group effect. There is some agreement in the literature that smaller classes may be beneficial (see Krueger 1999 and Angrist and Lavy 1999). In the present case, the older class is on average smaller, so that – if anything – this may lead to a downward bias of the true peer group effect on student outcomes. Using the estimated class size effects from Project STAR in Krueger 1999 as benchmark – if one is indeed willing to extend the results from Project Star to the current setting – the potential bias from the class size differences is about 0.09 standard deviations, which would indicate a reduction of the effect of being in the older class by about 20%.²³

Table 5: Teacher and student perception of learning environment

Panel A: Teacher perception			
Disciplinary problems with students	0.139*	(0.078)	559
Fraction of planned curriculum taught	-0.040***	(0.013)	561
Rate of students expected to finish primary school	-0.057***	(0.018)	562
Rate of students expected to finish secondary school	-0.060**	(0.025)	562
Panel B: Student perception			
Fellow students are noisy and disruptive	0.032***	(0.011)	13,630
Fellow students leave classroom early	0.050***	(0.011)	13,509
Fellow students learn taught material	-0.024***	(0.009)	13,469
Fellow students pay attention in class	-0.011	(0.008)	13,630
Teacher enforces student attention	-0.006	(0.005)	13,731
Teacher corrects homework	-0.014***	(0.005)	13,506
Teacher availability to clarify doubts	-0.027***	(0.007)	13,817
Teacher explains until all students understand	-0.023***	(0.007)	13,783
Teacher gives opportunity to express oneself	-0.025***	(0.007)	13,729
Teacher helps more some students	0.053***	(0.011)	13,480
Teacher interested in learning progress	-0.019***	(0.005)	13,775
Teacher needs to wait to start teaching	0.036***	(0.012)	13,630
Teacher absenteeism	0.026***	(0.009)	13,469

Notes: Entries are separate OLS estimates of the class rank on the perception of teachers and students of the teaching and learning environment in class. For each variable a separate regression has been estimated. The variables in the top panel are from the teacher questionnaire. The variable *disciplinary problems with students* is a dummy taking a value 1 if teachers report that there are problems with the discipline of students. The variables from the bottom two panels come from the student questionnaire of PROEB 2007. The variables have been recoded from categories ranging from “totally disagree” to “totally agree” on a scale from 0-1. All regressions control for school fixed effects and the full set of controls as in column (3) of Table 3. Heteroskedasticity consistent standard errors, clustered on the school level, are reported in parentheses. *, ** and *** denote significance at the 10%, 5% and 1% level, respectively.

²³ This is calculated as the difference in class size between the two classes, divided by the average class size difference in Project Star multiplied with the estimated effect of class size on standardized test scores (3/7.5*0.22 S.D.).

Indirect effects: behavioural responses of teachers and students

Despite the fact that teachers are observationally equivalent across classes, their teaching practices may differ as a response to the composition and behaviour of students in the class. To develop an understanding of the teacher's perception of the teaching environment they face in classes with a different composition of students, I use information from the teacher questionnaire of PROEB and regress an indicator for disciplinary problems on class rank (while controlling for the set of teacher controls as in column (3) of Table 3).²⁴ In Table 5, I find that teachers in the older classes report more likely that there are disciplinary problems with the students in the class (marginally significant at 10% level). It also seems that teaching is less efficient in these classes evidenced by the difference in the fraction of the curriculum taught (-0.04). Overall, teachers are also less confident in the competence of students in the older class. Teachers expect the rate of students completing primary school in the older class to be lower (by about 6%) compared to students in the younger class. The rate expected to complete secondary schools differs in a similar magnitude across classes.

The learning environment is also perceived to be different by students in these classes. I use information from the student questionnaire on items related to the behaviour of their peers and teaching practices to learn about the learning environment. The responses that express agreement with different statements range from 0 to 1 and I regress these responses on the class rank and the full set of student and teacher controls as in column (3) of Table 3. The results are reported in panel B of Table 5.

Students in older classes more often report that their classmates are noisy and disruptive (0.032), which is a 6% difference compared to the mean. The probability of students leaving class early is substantially higher in the older classes (0.050, a 19% difference), which may contribute to the disruption of teaching in these classes. The less favourable learning

²⁴ The summary statistics of the variables can be found in Table A4.

environment is also confirmed by students in the older class reporting more often that their teacher needed to wait to start teaching at the beginning of class due to noise (0.036, a 6% difference).

The composition and behaviour of students may also lead to teachers adjusting their teaching practices. Students in the older class report that their teacher is available less to clarify doubts about the class material. The coefficient is -0.027 and statistically significant at the 1% level, which is 34% of a standard deviation of the mean. Similarly, students in the older class feel that the opportunity to express their opinion in class is substantially lower (-0.025, which is about 25% of a standard deviation of the mean). Further evidence of an effect on teaching practices through the impact on the distribution of instruction time is given by the difference in the answers on whether the class teacher helps some students more than other students. The estimate for this variable shows a 0.053 difference between classes. It appears that teachers in the older class are compelled to distribute their attention and instructional time more unequally, possibly devoting relatively more time to specific groups of students or addressing the same material again, but targeting it at different skills levels within the same class. With more heterogeneous groups, teachers may be less able to teach to the median student and they may need to specifically address the needs of students at the tails of the distribution. The distributional features of the class composition also possibly result in teachers being less able to devote enough time until every student has comprehended the material (-0.023, which is about 27% of a standard deviation of the mean). The higher dispersion in age and ability possibly demands that teachers address different skill levels separately, contributing to the difference in the fraction of the curriculum completed across the two sets of classes.

The less favourable teaching environment may also have an effect on teacher motivation. Students of the older class report more often (0.026, an 11% difference to the mean) that a teacher had been absent from school. The effect on absence of teachers may be interpreted as a response to the more deprived teaching environment. In turn, although difficult to quantify in terms of hours of instruction lost, teacher absence may also affect the achievement of students, creating

negative feedback effects between class composition, teacher and student behaviour. Teachers also appear to show less of an interest in the learning of their students and are less likely to mark their homework, all possibly contributing to the worse learning environment in the older class. These differences in teaching practices are particularly striking, given that I do not find any differences in any of the observable characteristics of teachers in Table 4.

These results are in line with the findings of Lavy, Paserman and Schlosser (2012), which show that a higher proportion of low ability students has a detrimental impact on teaching practices of teachers, lead to more classroom disruption, and worse student-student and student-teacher interaction.

Table 4 also shows that the percentage of students who do not participate in the PROEB test, due to illness or other reasons, differs between the two classes. Although the non-response rate differs between younger and older classes for the peer group and is about 9% higher in the older classes, the non-response rate has a smooth transition across the discontinuity point. The size of the RD estimate for the non-participation rate at the threshold is very small and not statistically significant, so that the estimates are very unlikely confounded by the differential non-response rate of students on either side of the cut-off point.²⁵

Opening the black box of the peer-group effect: heterogeneous treatment across schools

To acquire some understanding of the distribution of effects across schools, I estimated school-specific discontinuities in maths test scores. As differences of mean peer variables between classes differ across schools, treatment also differs in respect of the composition of the peer class environment. Figure A3 plots the kernel density estimates of the school-specific discontinuities and shows the relatively symmetric distribution of effects.

In the previous sections, the different potential channels through which the peer composition in this setting may lead to the estimated drop in academic performance close to the cut-off point, have been introduced. Subsequently I aim at quantifying the contribution of a number of key

²⁵ The data appendix provides information on how the non-response rate on the class level and around the threshold has been established.

differences across classes to the estimated group effects. For this purpose, I make use of the setup at hand, with discontinuities in the 363 schools, which allow examining the role of different observable characteristics of the peer group in explaining the gap in academic achievement. More precisely, the fact that the difference in the characteristics of peers between children in younger and older classes differs across schools can be used to gain some understanding of the role of the underlying potential channels. For students around the cut-off point, class characteristics, such as the socioeconomic composition of their peer group, are arguably quasi-random, and the difference in these characteristics between classes varies across schools can be related to the size of the test score difference across classes at the threshold.

For this purpose, I use a two-stage minimum-distance estimator, where in a first stage I estimate the size of the discontinuity in test scores at the cut-off and the differences in peer characteristics between the two classes by 2SLS separately for each school.²⁶ In the second stage, the estimated discontinuities in test scores are used as dependent variables and are regressed on the estimated differences in class characteristics z_{cs}

$$b_s = \alpha_0 + \alpha_1 \Delta z_{cs} + u_s \quad (3)$$

where b_s are the estimated discontinuities in test scores for marginal students from the first stage.

Because the estimates of b_s are based on regressions using individual data, the minimum distance estimator is derived by minimising the weighted difference between the auxiliary parameters from the first stage estimation, where the weights are equal to the reciprocal of the square of the standard errors of the first stage running minimum-distance weighted least squares.²⁷ I also include school and teacher level characteristics as controls in (3).

Obviously, to the extent that there are other unobservable class level characteristics that affect outcomes and are correlated with the included regressors, the minimum distance estimates will

²⁶ Wolfowitz (1957) introduced the minimum-distance estimator. See Kodde et al. (1990) for details.

²⁷ Because the explanatory variables are estimated from a first-stage procedure, generally the standard errors and test statistics may be invalid because they ignore the sampling variation of the estimated regressors. There is nevertheless one exception, as in this case, when testing the null hypothesis $H_0: \alpha_1 = 0$, the test statistics has a limiting standard normal distribution, so that no adjustment of the standard errors is required in this instance (Wooldridge 2010). This holds under a usual homoscedasticity assumption. The heteroskedasticity-robust statistic is valid if heteroskedasticity is present under the null and I therefor report robust standard errors in Table 6.

confound the effect of such variables with the effect of the included regressors. For example, if being older is also associated with lower innate ability, for example, because older students have previously repeated a grade, but I am unable to measure innate ability, the measure of the average age of peers will also pick up the effect of having less able peers. It is, consequently, not possible to disentangle the effect of ability heterogeneity from the effect of age heterogeneity in this context. In addition, many of the peer characteristics are highly correlated and including them all as explanatory variables may lead to multicollinearity in (3). To address potential multicollinearity and because I am interested in the overall effect of exogenous peer characteristics I summarize all available socio-economic variables in an SES index using Principal Component Analysis.²⁸ I am then particularly interested in the effect the difference in age dispersion, mean age, mean grade repeated and class size have on the estimated math performance gap, in addition to the measure of socio-economic status.

Table 6 provides the coefficients of the above two-stage procedure.²⁹ Column (1) reports the effects for all of these explanatory variables, columns (2) - (6) when entering the regressor one-by-one to test for the role of multicollinearity. All specifications control for teacher and school characteristics. Out of all the regressors, only age dispersion is significant and contributes positively to the gap in math test scores. A one-month difference in the standard deviation of age explains about 0.033 of standard deviation in maths test scores, which is just under 8% of the estimated discontinuity. Mean age, mean grades repeated and class size do not have the expected sign, but have very large standard errors and are not significant at any conventional level of significance. The SES index has the expected sign, but is not significant in the multivariate regression. In column (2) where I include only age dispersion with the controls the coefficient is

²⁸ I included the estimated discontinuities in sex, white, mixed, black, Asian, indigenous students, fraction of HH with maids, Bolsa Família, number of bathrooms, books, cars, computers, fridges, freezers, radios, washing machines, dryer, DVD players, TV sets, video players in the PCA analysis and high values of the Kaiser–Meyer–Olkin measure indicate ($>.80$) indicate that all the variables are adequate for inclusion on the SES index. For each of these variables the unexplained variance is low, pointing to the high correlation between these variables. The first principal component explains 56% of the total variance.

²⁹ The dependent variable of the test score gap carries a positive sign, so that a larger positive value refers to a larger negative discontinuity in maths test scores between the two classes.

essentially unchanged. In columns (3) to (6) I include the other variables one-by-one, and only the coefficient for the SES index is marginally significant and larger than in the multivariate regression, pointing to a remaining potential role for multicollinearity.³⁰ Although the results from this exercise should be considered with caution regarding a causal interpretation, they point to an important role of the age dispersion for explaining the gap in math test scores across the class discontinuity. Together with the results on behavioural responses by teachers and students, the findings draw a picture on the potential effect of the more dispersed age distribution in the older classes on the performance of students: The more heterogeneous classes may crucially contribute to the differences in teaching practices shown above, including teachers being less able to spend equal time on all students in the more heterogeneous classes. Similarly, student's may respond to the more heterogeneous class composition and the teaching response by teachers and some students may find themselves idle while teachers address subsets of students in the class, contributing to a less efficient learning environment.³¹

VIII. Conclusions

In this paper, I use an RD design that exploits the rule, which assigns students of a given cohort to classes according to their ranking along the age distribution to estimate the effect of group membership on standardised maths test scores. The RD design allows us to compare students who are very similar in age but find themselves being assigned to classes with either younger or older students. By exploiting this rule, I provide evidence of strong negative effects on maths achievement for marginal students being in a class with older peers. I find that marginal students who are assigned to the older classes have maths test scores that are about 40% of a standard deviation lower than those of students assigned to the younger classes. While there is

³⁰ I have also estimated models where I included all the individual peer characteristics in (1), summarized in the SES index. All the coefficients in these regressions are imprecise, probably due to considerable correlation between these variables.

³¹ These findings are in line with the results of Hoxby and Weingarth's study (2006) on the importance of the age dispersion in the reference group on academic achievement.

Table 6: Treatment effects across schools

Difference in class means	(1)	(2)	(3)	(4)	(5)	(6)
Age dispersion	3.330** (1.481)	3.016*** (1.109)				
Mean age	-2.085 (1.523)		-0.303 (1.100)			
Mean grades repeated	-5.754 (50.827)			-40.598 (51.656)		
Class size	-1.430 (2.722)				-1.428 (1.133)	
SES index	-1.430 (2.722)					-4.339* (2.468)
Teacher and school controls:	yes	yes	yes	yes	yes	yes
Number of observations:	363	363	363	363	363	363
R ²	0.368	0.350	0.324	0.328	0.337	0.330

Notes: The dependent variable is a measure of the absolute size of the discontinuity in math test scores at the cut-off point at the school level estimated by 2SLS. The entries report coefficients from the second stage of the minimum distance estimation, where weights are equal to the inverse of the standard errors of the estimates of the first stage. Independent variables are the discontinuities of peer values the age distribution, mean age, a measure for repetition and an index for socioeconomic status estimated by 2SLS. The SES index was derived using Principal Component Analysis on 19 variables (the estimated discontinuities in sex, white, mixed, black, Asian, indigenous students, fraction of HH with maids, Bolsa Família, number of bathrooms, books, cars, computers, fridges, freezers, radios, washing machines, dryer, DVD players, TV sets, video players). All regressions control for teacher characteristics school characteristics (teacher age, teacher experience, teacher education, teacher seniority, measures of quality classrooms, number of school computers, quality of school books, number of school books, broadband access and teaching material. Heteroskedasticity robust standard errors are reported in parenthesis. *, ** and *** denote significance at the 10%, 5% and 1% level, respectively.

no evidence for common shocks in the form of differences in teacher quality driving these estimates, I show that the peer composition differs substantially across the two set of classes. Older classes are composed of students who are on average more likely to be male, from lower socio-economic households and with a higher fraction of black and mixed background. The classes have a much higher fraction of repeaters and have a much more dispersed age distribution. Using variation in class composition from more than 350 school discontinuities, I present some suggestive evidence that differences in the age distribution may play a crucial role for explaining the large negative effect on test scores of being in the older class. The difference in mean age, the number of repeaters and class size do not have a statistically significant effect on the math test score gap. There is some evidence for a potential role of socio-economic status to play a role, but the effect does not hold in multivariate regression, possibly due to multicollinearity. The evidence in favour of a role of the age distribution may help explain the differences in observed teaching practices. Teachers in the older classes are – according to students – less likely to distribute their attention equally among students in the class, they are less likely to clarify doubts of students regarding the content and they are less likely to explain until all students understand the content. These differences are striking because I find no evidence in favour of any differences in pre-determined teacher characteristics, which may be indicative of systematic sorting of teachers. Students also differ in their behaviour and are reported to be noisier and more disruptive in the older class and are more likely to leave classroom early, contributing to the adverse learning environment in the older classes, possibly also in response to the difference in the student composition and the teaching practices. These results fit an interpretation where class heterogeneity, in age or potentially in related other characteristics such as the heterogeneity in ability, contributes to a learning environment that is substantially different across classes and which may explain the observed differences in teaching practices and in the behavioural responses of students documented in this paper. These findings also contribute to an emerging part in the peer effects literature taking that explicitly considers

group heterogeneity as relevant factor for estimating peer effects (De Giorgi, Pellizzari and Woolston 2010).

The paper also contributes to some extent to the literature on relative age effects in education. Concurrently with being in different peer environments, marginal students are also either the oldest or the youngest in their respective classes and, apart from the effect from being assigned to classes with different peer characteristics and their distribution, there could be a separate pure relative age effect at work. It is, nevertheless, debatable whether conceptually there is a difference between a potential pure relative age effect and an age peer group effect, and, given the identification strategy, these effects would be practically indistinguishable. Moreover, there is mixed evidence on the existence of a separate pure relative age effect in the literature.³²

³² Elder and Lubotsky (2009) show that a commonly postulated positive relationship between achievement and school entry age is primarily driven by the skills older children acquired prior to kindergarten rather than absolute or relative age effects. Using experimental data from Project STAR, Cascio and Whitmore Schanzenbach (2016) find some small positive effects of having older children in the classroom conditional on one's own age, which is contrary to findings in this paper. Crawford, Dearden and Meghir (2010) find that the month of birth matters in national achievement tests in England, and even show long-run effects beyond post-compulsory education. As the identification strategy employed in this paper is based on the discontinuity around the median age in the cohort, the estimated effects are not confounded by relative age effects at the extremes of the age distribution, that is, being the youngest or oldest in the cohort, so that targeting the curriculum to a specific age group will not bias the estimated effects. There exists a related literature that looks at the rank in the distribution more generally providing evidence on the importance of the relative rank position apart from age (Murphy and Weinhardt 2014, Elsner and Ipshording 2016).

References

- Adams-Byers, J., Whitsell, S. and Moon, S. (2004), Academic and Social/Emotional Effects of Homogeneous and Heterogeneous Grouping, *Gifted Child Quarterly* 48, 7–20.
- Ammermueller, A. and Pischke, J. (2009), Peer Effects in European Primary Schools: Evidence From the Progress in International Reading Literacy Study, *Journal of Labor Economics*, 27, 315–348.
- Angrist, J. and Krueger, A. (1991), Does Compulsory School Attendance Affect Education and Earnings?, *Quarterly Journal of Economics* 106, 979–1014.
- Angrist, J. and Lang, K. (2004), Does School Integration Generate Peer Effects? Evidence from Boston's Metco Program, *American Economic Review*, 94, 1613–1634.
- Angrist, J. and Lavy, V. (1999), Using Maimonides' Rule to Estimate the Effect of Class Size on Children's Academic Achievement, *Quarterly Journal of Economics*, 114, 533–575.
- Bandiera, O., Barankay, I. and Rasul, I. (2010), Social Incentives in the Workplace, *Review of Economic Studies*, 77, 1047–1094.
- Bayer, P., Hjalmarsson, R. and Pozen, D. (2009), Building Criminal Capital Behind Bars: Peer Effects in Juvenile Corrections, *Quarterly Journal of Economics*, 124, 105–147.
- Betts, J. and Shkolnik, J. (1999), The Effects of Ability Grouping on Student Achievement and Resource Allocation in Secondary Schools, *Economics of Education Review*, 19, 1–15.
- Crawford, C., Dearden, L. and Meghir, C. (2010), When you are Born Matters: The Impact of Date of Birth on Educational Outcomes in England, *DoQSS Working Paper No. 10-09*.
- Carrell, S. and Hoekstra, M. (2010), Externalities in the Classroom: How Children Exposed to Domestic Violence Affect Everyone's Kids, *American Economic Journal: Applied Economics*, 2, 211–228.
- Cascio, E. and Whitmore Schanzenbach, D. (2016), First in the Class? Age and the Education Production Function, *Education Finance and Policy*, forthcoming.
- De Giorgi, G., Pellizzari, M. and Woolston, W. (2012), Class Size and Class Heterogeneity, *Journal of the European Economic Association*, 10, 795–830.
- Duflo, E., Dupas, P. and Kremer, M. (2011), Peer Effects, Teacher Incentives, and the Impact of Tracking: Evidence from a Randomized Evaluation in Kenya, *American Economic Review*, 101, 1739–1774.
- Elder, T. and Lubotsky, D. (2009), Kindergarten Entrance Age and Children's Achievement: Impacts of State Policies, Family Background, and Peers, *Journal of Human Resources*, 44, 641–683.

- Elsner, B. and Ipschording, I. (2016), A Big Fish in a Small Pond: Ability Rank and Human Capital Investment, *Journal of Labor Economics*, forthcoming.
- Figlio, D. and Page, M. (2002), School Choice and the Distributional Effects of Ability Tracking: Does Separation Increase Inequality? , *Journal of Urban Economics*, 51, 497–514.
- Gomes-Neto, J. and Hanushek, E. (1994), Causes and Consequences of Grade Repetition: Evidence from Brazil, *Economic Development and Cultural Change*, 43, 117–148.
- Hanushek, E., Kain, J., Markman, M. and Rivkin, S. (2003), Does Peer Ability Affect Student Achievement? , *Journal of Applied Econometrics*, 18, 527–544.
- Hoxby, C. (2000), Peer Effects in the Classroom: Learning from Gender and Race Variation, *NBER Working Paper 7867*.
- Hoxby, C. and Weingarth, G. (2006), Taking Race out of the Equation: School Reassignment and the Structure of Peer Effects, *Unpublished Manuscript*.
- Imbens, G. and Lemieux, T. (2007), Regression Discontinuity Designs: a Guide to Practice, *Journal of Econometrics*, 142, 615–635.
- Kodde, D., Palm F. and Pfann, G. (1990), Asymptotic Least-squares Estimation Efficiency Considerations and Applications, *Journal of Applied Econometrics*, 5, 229–43.
- Kremer, M. (1997), How Much does Sorting Increase Inequality, *Quarterly Journal of Economics*, 112, 115–139.
- Krueger, A. (1999), Experimental Estimates of Education Production Functions, *Quarterly Journal of Economics*, 114, 497–532.
- Lavy, V., Paserman, D. and Schlosser, A. (2012), Inside the Black Box of Ability Peer Effects: Evidence from Variation in Low Achievers in the Classroom, *Economic Journal*, 122, 208–237.
- Lavy, V. and Schlosser, A. (2011), Mechanisms and Impacts of Gender Peer Effects at School, *American Economic Journal: Applied Economics*, 3, 1–33.
- Lavy V., Silva O. and Weinhardt, F. (2012), The Good, the Bad and the Average: Evidence on the Scale and Nature of Ability Peer Effects in Schools, *Journal of Labour Economics*, 30, 367–414.
- Lee, D. and Lemieux, T. (2010), Regression Discontinuity Designs in Economics, *Journal of Economic Literature*, 48, 281–355.
- Lindert, K., Linder, A., Hobbs, J. and de la Brière, B. (2007), The Nuts and Bolts of Brazil's Bolsa Família Program: Implementing Conditional Cash Transfers in a Decentralized Context, *Social Protection Discussion Paper 0709*, World Bank.

- Lyle, D. (2009), The Effects of Peer Group Heterogeneity on the Production of Human Capital at West Point, *American Economic Journal: Applied Economics*, 1, 69–84.
- Mas, A. and Moretti, E. (2009), Peers at Work, *American Economic Review*, 99, 112–145.
- Manski, C. (1993), Identification of Endogenous Social Effects: the Reflection Problem, *Review of Economic Studies*, 60, 531–542.
- McCrary, J. (2008), Manipulation of the Running Variable in the Regression Discontinuity Design: a Density Test, *Journal of Econometrics*, 142, 698–714.
- Ministry of Education. (2004), *Ensino Fundamental de Nove Anos – Orientações Gerais*, Secretariat of Basic Education, Federal Brazilian Ministry of Education. Brasília.
- Murphy, R. and Weinhardt, F. (2014), Top of the Class: the Importance of Ordinal Rank, CESifo Working Paper No. 4815.
- Patrinos, H. and Psacharopoulos, G. (1996), Socioeconomic and Ethnic Determinants of Age-grade Distortion in Bolivian and Guatemalan Primary Schools, *International Journal of Educational Development*, 16, 698–714.
- Robinson, J. (2008), Evidence of a Differential Effect of Ability Grouping on the Reading Achievement Growth of Language-minority Hispanics, *Educational Evaluation and Policy Analysis*, 30, 141–180.
- Sacerdote, B. (2003), Peer Effects with Random Assignment: Results for Dartmouth Roommates, *Quarterly Journal of Economics*, 116, 118–136.
- Urquiola, M. (2006), Identifying Class Size Effects in Developing Countries: Evidence from Rural Bolivia, *Review of Economics and Statistics*, 88, 171–177.
- Van der Klaauw, W. (2002), Estimating the Effect of Financial Aid Offers on College Enrolment: A Regression-discontinuity Approach, *International Economic Review*, 43, 1249–1287.
- Whitmore, D. (2005), Resource and Peer Impacts on Girls’ Academic Achievement: Evidence from a Randomized Experiment, *American Economic Review*, 95, 199–203.
- Wolfowitz, J. (1957), The Minimum Distance Method, *The Annals of Mathematical Statistics*, 28, 75–88.
- Wooldridge, J. (2010), *Econometric Analysis of Cross Section and Panel Data*, MIT Press, Cambridge, Massachusetts.
- Zimmer, R. (2003), A New Twist in the Educational Tracking Debate, *Economics of Education Review*, 22, 307–315.
- Zimmerman, D. (2003), Peer Effects in Academic Outcomes: Evidence from a Natural Experiment, *Review of Economics and Statistics*, 85, 9–23.