DISCUSSION PAPER SERIES

DP16471

Let's stay together: the effects of repeated student-teacher matches on academic achievement

Facundo Albornoz , David Contreras and Richard Upward

PUBLIC ECONOMICS

ISSN 0265-8003

Let's stay together: the effects of repeated student-teacher matches on academic achievement

Facundo Albornoz , David Contreras and Richard Upward

Discussion Paper DP16471
Published 19 August 2021
Submitted 14 August 2021

Centre for Economic Policy Research
33 Great Sutton Street, London EC1V 0DX, UK
Tel: +44 (0)20 7183 8801
www.cepr.org

Let's stay together: the effects of repeated student-teacher matches on academic achievement

Abstract

We explore the effectiveness of repeating the student-teacher match on test scores, for the universe of 8th graders in Chile using information on all student-teacher matches across multiple subjects and years, and a national, anonymous measure of test scores. Also, we exploit a plausibly exogenous source of variation in the process of repeating matches generated by a discontinuity in teacher retention at the legal retirement age. Repeating matches has a robust positive effect on test scores which aggregates up to the student, class, and school-level. As channels, we report a positive effect on attendance, progression, student behaviour and teacher expectations.

JEL Classification: I20, I21, I28

Keywords: N/A

Facundo Albornoz - facundo.albornoz@nottingham.ac.uk
University of Nottingham and CEPR

David Contreras - david.contrerasgomez@nottingham.ac.uk
University of Nottingham

Richard Upward - richard.upward@nottingham.ac.uk
University of Nottingham

Acknowledgements
We thank Abhijeet Singh, Antonio Cabrales, Gianni de Fraja, Miguel Urquiola and participants at the GEP/CEPR Education Economics Workshop and the CESifo Area Conference on Economics of Education.

Let's stay together: the effects of repeated student-teacher matches on academic achievement*

Facundo Albornoz†

University of Nottingham and CEPR

David Contreras‡

University of Nottingham

Richard Upward§

University of Nottingham

August 14, 2021

*We thank Abhijeet Singh, Antonio Cabrales, Gianni de Fraja, Miguel Urquiola and participants at the GEP/CEPR Education Economics Workshop and the CESifo Area Conference on Economics of Education.

†facundo.albornoz@nottingham.ac.uk. Address: Room B66a Sir Clive Granger Building, University Park Nottingham, UK. Phone: +44 115 95 15474.

‡david.contrerasgomez@nottingham.ac.uk. Address: Room B49 Sir Clive Granger Building, University Park Nottingham, UK. Phone: +44 755 146 1260.

§Corresponding author. richard.upward@nottingham.ac.uk. Address: Room B3 Sir Clive Granger Building, University Park Nottingham, UK. Phone: +44 115 95 14735.

Abstract

We explore the effectiveness of repeating the student-teacher match on test scores, for the universe of 8th graders in Chile using information on all student-teacher matches across multiple subjects and years, and a national, anonymous measure of test scores. Also, we exploit a plausibly exogenous source of variation in the process of repeating matches generated by a discontinuity in teacher retention at the legal retirement age. Repeating matches has a robust positive effect on test scores which aggregates up to the student, class, and school-level. As channels, we report a positive effect on attendance, progression, student behaviour and teacher expectations.

JEL classification: I21, I25

Keywords: student-teacher matches, student achievement, looping

1 Introduction

Each year, school managers must allocate teachers to groups of students. Consider a school with two maths teachers, and two groups of students who progress from grade 7 to grade 8. Each teacher could specialise in a particular grade: teacher 1 takes both groups in grade 7, and teacher 2 takes both groups in grade 8. Under this allocation, all students are matched with a new teacher in grade 8. An alternative arrangement is to repeat the student-teacher match, which is called "looping" in the educational literature. Under this allocation, each teacher is assigned to a single group of students which they teach in both grade 7 and 8. Students who remain in the same group between grades will be matched with the same teacher in both grades. Students who change group between grade 7 and 8 will be matched with a new teacher, but will typically still be in a group in which most students have the same teacher in both grades. Does looping have any impact on student achievement? If yes, how and through which mechanisms? This paper attempts to provide answers to these questions.

Understanding the effect of looping is important for at least two fundamental reasons. First, it is widely used in some school systems. Although systematic quantitative evidence on the prevalence of looping does not appear to be available, it seems to be widespread in German elementary schools (Zahorik & Dichanz 1994), in Chinese schools at all levels (Liu 1997) as well as in Finland, Japan, Sweden, Israel and Italy (Tourigny, Plante & Raby 2019). In the case we study, Chile, over 50% of students progressing from year 7 to 8 have the same teacher in both grades. Thus, measuring the effect of looping-based teacher-student allocations on student outcomes is potentially of great importance. Second, repeating student-teacher matches only requires a re-assignment of existing teaching resources without significant additional costs. Thus, if it works, looping can be a budget-neutral way to improve student achievement.

In this paper, we use rich, comprehensive student-teacher data to explore the effect of repeating the student-teacher match on students' test scores for 8th graders in Chile.

Unusually, we have information on all student-teacher matches across multiple subjects and multiple years, and we have a national, anonymous measure of student test scores which is uncontaminated by any teacher or school biases in grading. However, even with these data, estimating the causal effect of repeating the student-teacher match is challenging for two reasons. First, because of non-random selection into repeat matches. Looping may be more common for certain types of school, teacher or student. Furthermore, student-teacher matches which are successful in one year may be more likely to be repeated. Second, even if one could randomly allocate repeat matches, those matches will tend to have more experienced teachers. This arises because, in order to repeat a match, the teacher must have taught at the same school in the previous year, while new matches are drawn from a pool which includes teachers who are recently hired.

We control for selection by schools, teachers or students by exploiting within-school, within-student and within-teacher variation in repeat matches which occurs across subjects and across time. Further, because the same teachers are observed in multiple classes in the same year, we can make within-teacher-by-year comparisons to control for the resulting experience gap. Using these fixed-effect methods we find that repeating a match increases student performance by about 0.02 standard deviations. This is equivalent to the effect of improving teacher quality by 0.1–0.2 standard deviations.[1] A value-added specification yields similar results.

However, fixed effects and value-added methods do not fully mitigate the concern that school managers (or teachers) might decide to repeat matches based on the performance of existing matches. However, we are able to show that residual performance measures from the earlier grades have almost no explanatory power for the formation of new matches. We also utilise a situation in which the teacher-student match is broken for exogenous reasons, namely the discontinuity in repeat matches which occurs when teachers reach the legal retirement age (LRA). Effectively, we compare the performance of grade 8 students whose grade 7 teacher reached the LRA in the previous year with

[1] Using estimates from Rivkin, Hanushek & Kain (2005) and Rockoff (2004).

grade 8 students whose grade 7 teacher reaches the LRA in the current year. Grade 8 students whose grade 7 teacher reached the LRA in the previous year are far more likely to be allocated a new teacher, and hence are far less likely to experience a repeat match. The discontinuity arises because of small differences in the date of birth of different grade 7 teachers. Using this discontinuity design, we obtain larger but more imprecise estimates of the benefit of repeating student-teacher matches. These estimates still allow us to reject the null of no effect and suggest that our fixed-effects estimates are not biased upwards by positive selection.

We then investigate whether the effects of repeat matches aggregate up to the student, class or subject level. The positive effects we observe at the student-subject level may be misleading if there is substitution of a fixed amount of effort by each student towards subjects with familiar teachers, at the expense of subjects with new teachers. We therefore test whether the positive effects of repeat matches aggregate up to the student, class and school level. Reassuringly, we find student, class and school-level estimates are all slightly larger than the equivalent student-subject level estimates.

We also explore several potential channels through which looping may improve student outcomes. Using evidence from a survey of teachers, we assess the effect of repeat matches on the learning environment at the class level. Educational research has emphasised the positive relationship between school effectiveness and a co-operative school environment. The literature has shown that a positive and sustained school climate[2] is correlated with higher levels of students' motivation and engagement, school attendance, graduation rates and teacher retention (Thapa, Cohen, Guffey & Higgins-D'Alessandro 2013). In addition, recent studies (Bryk, Sebring, Allensworth, Easton & Luppescu 2010, Kraft, Marinell & Shen-Wei Yee 2016, Klugman 2017) have established a positive causal impact of school climate on students achievement on standardised test scores. We find that that in classes with more student-teacher matches, students have

[2]The school climate reflects the quality of the relations between the members of the educational community. For a comprehensive review on school climate literature, see Thapa, Cohen, Guffey & Higgins-D'Alessandro (2013).

higher attendance, teachers report better classroom behaviour and have higher expectations of their students' academic potential. Our finding that the student-subject level effect is slightly smaller than the student, class, or school-level effect is consistent with the notion that greater student-teacher familiarity has a positive effect on school climate which may affect outcomes of students who do not themselves repeat the match, such as those who join a classroom where the other students are looping.

Very few papers have attempted to formally evaluate the effectiveness of repeat matches for student achievement. An exception is Hill & Jones (2018), who assess the impact of repeat matches on the academic achievement in elementary public schools from North Carolina using a similar fixed effects approach. Their estimated effect on test scores is positive, significant and similar to ours. We build on their findings by exploiting data on the universe of Chilean students and teachers over a longer period and present the first evidence that the positive effect of looping is robust to relaxing the assumption of no selection on match quality. We show that looping is a successful strategy for older children in a setting in which looping is common, and we show that plausible mechanisms exist.

Repeating student-teacher matches necessarily implies greater student-teacher familiarity. In this sense, our analysis is related to Fryer (2018), who investigates the effect of teacher specialisation by subject, and finds that specialisation decreases students' achievement and attendance, and increases student behaviour problems. Fryer suggests that these findings could be explained by the decrease in interactions between teachers and students, caused by teachers' subject specialisation. Our findings support this view in a different context, from a different policy, and provides complementary evidence on how student-teacher familiarity manifests in better classroom behaviour.

A recent literature emphasises complementarities between teacher and student characteristics (e.g. Aucejo, Coate, Fruehwirth, Kelly & Mozenter 2018, Graham, Ridder, Thiemann & Zamarro 2020). This implies that improving teaching-to-classroom assignments may lead to better student outcomes. Graham, Ridder, Thiemann & Zamarro (2020) experiment with different assignments to show that overall achievement in elementary schools in the US can increase by at around 0.02 standard deviations without changes in existing teaching resources. Of course, a precise performance-improving assignment of teachers to classrooms requires information that it is not necessarily available for school managers. Our paper complements these findings by providing a simple and feasible assignment rule that delivers results which are at least as large, if not larger.

A number of qualitative and small-scale quantitative studies in the educational literature have investigated the effectiveness of looping, including Bogart (2002), Nichols & Nichols (2002), Cistone & Shneyderman (2004), Tucker (2006) and Franz, Thompson, Fuller, Hare, Miller & Walker (2010). Cistone & Shneyderman note that looping is widespread in primary schools in certain countries, including Germany and Japan, but rarely used in others. Most of these studies consider elementary schools: Kerr (2002) stresses that very few studies consider effects on older children. These studies overwhelmingly argue that looping improves student outcomes. For example, Cistone & Shneyderman (2004) find that looping improved student attendance and increased the rate at which students progressed successfully to the next grade. It is commonly suggested that looping has these positive benefits because it saves considerable time at the start of the new school year. Cistone & Shneyderman (2004) argue that looping "allows teachers to save time at the beginning of the second year of the loop by making unnecessary the usual transitional period typically spent on getting acquainted with new students as well as setting classroom rules, expectations, and standards." The same idea is also argued by Burke (1996), Little & Dacus (1999) and Black (2000). A teacher cited by Little & Dacus (1999, p.43) explains: "Gone were the lectures about daily procedures and classroom rules. Gone were the weeks of testing, trying to determine a student's reading level. The teachers and students started the year with a bang and ended further along than the teachers had anticipated." The literature also argues that looping allows teachers to build closer relationships with the students and parents, along with a better understanding of the strengths, weaknesses and personalities of their students. Looping also allows teachers to implement a smooth transition across grade levels and develop a more cohesive curriculum.

The literature also recognises that looping may also have disadvantages. First teachers may find it more difficult to teach a multi-year rather than single-year curriculum. Second, teachers may lose grade-specific human capital, which Ost (2014) finds contributes up to one-third as much as general teaching experience, at least for maths scores. Finally, even if repeated matches are more efficient, they may also increase inequality in student outcomes, because, as noted by Bogart (2002), some unlucky students will spend two or more years with an ineffective teacher. Assigning students to new teachers each year mitigates these inequality concerns.

This educational literature provides useful insights on how looping may affect the learning process, but does not provide a systematic assessment of its overall causal effect. Our paper is a contribution in that direction, and we show that the benefits of looping outweigh the costs, at least on average.

The remainder of the paper is organised as follows. Section 2 describes our data and the relevant institutional features of the Chilean school system. Section 3 explains the econometric framework and estimates the effect of repeated student-teacher matches at the student-subject level. We begin with fixed-effects methods which maintain the assumption that selection into repeat-matches is exogenous to the quality of existing matches. We then relax this assumption by exploiting the discontinuity at the LRA as a source of exogenous variation in repeat match formation. In Section 4 we estimate the effects of repeated matches at the student, class and school level, which may be more informative as to the effectiveness of a policy of repeating student-teacher matches, since there may be spillover or substitution effects within and between students. In Section 5 we report the results from large-scale teacher survey results which support the hypothesis that repeated matches improve behaviour in the classroom and raise teacher expectations of future student performance.

2 Data and institutional background

We use three different datasets provided by the Chilean Ministry of Education. First, we use the complete school enrolment records of all students in Chile from 2002 onwards. The database contains yearly information on the students enrolled in primary school (grade 1 to grade 8) and high school (grade 9 to grade 12). These records contain a consistent student ID, a school ID and a "class" ID. In Chilean schools, a class is a fixed group of students who take subjects together: every student in our sample is in the same group (class) in grade 8 for all four subjects we consider. The enrolment records include individual school grades (awarded by teachers) in each subject and the individual attendance rate. The grading system in Chile is 1 to 7 by increments of 0.1, and schools are free to set their own grading standards. To make school grades comparable, we standardise school grades at the school level.[3]

Second, we use comprehensive teachers' administrative records. These records contain information on teacher gender, age and experience. This database includes the same class ID as in the enrolment records, which allows us to associate each class of students in each subject with a teacher in each year. The enrolment records matched to the teacher records allow us to measure whether a student has the same teacher in a subject for successive years.

Third, we use data on students' achievement in *Sistema de Medición de la Calidad de la Educación* (SIMCE) tests. This is a standardised test administered by the Ministry of Education to all students in certain grades, and is the main instrument to measure the quality of education in Chile. The SIMCE is administrated by external examiners, and provides information about students' performance relative to the coun-

[3]We do not use these school grades as an outcome measure because they may reflect teacher biases as well as student performance (Contreras 2019).

try's National Curriculum Framework. We use standardised test scores for 8th graders in four years: 2004, 2007, 2009 and 2011, in four different subjects: Spanish, maths, social sciences and natural sciences.[4] In these three years, SIMCE tests were taken by 1,056,458 students, 97.8% of the students enrolled in 8th grade, covering 98.4% of schools in operation.[5]

The SIMCE data also contains information on school characteristics (including whether a school is public or private) and information from surveys of parents and teachers. The parents' survey provides information on family socio-economic background, including mother's schooling and monthly household income (banded). For years 2009 and 2011, the teachers' survey provides information about perception of classroom behaviour and the future performance of the class. Teachers complete a separate survey for each class they teach.

We therefore have information on students $i = 1 \ldots N$ who are observed in 8th grade in one of four different years ($t = 2004, 2007, 2009, 2011$). Each student has SIMCE test scores in four subjects $s = 1, 2, 3, 4$. Students are grouped together in classes c. A class-subject combination has a specific teacher j, school k and year t. We start with a sample of 789,270 students. After excluding observations without valid test scores, student or teacher characteristics, we are left with a sample of 696,482 students, 46,256 teachers, 31,837 classes and 6,260 schools. Overall, the estimation sample represents 76.3% of the students enrolled in 8th grade who took all the SIMCE tests. Information from teachers about classroom behaviour and future class performance is available for 9,498 classes for each of the four subjects.

A *repeat match* takes place when a student has the same teacher in the same subject

[4] We focus on grade 8 in these four years because we have information on all four subjects' SIMCE test scores, and we exploit the variation across subjects.

[5] The SIMCE test is not taken by students in special education or adult education. In addition, there are cases in which the test cannot be taken because schools are closed temporarily or because individuals students cannot attend. Cuesta, González & Philippi (2020) find that high-performing students are more likely to take the SIMCE test, and that the size of this effect varies across school. Our findings, however are based on a within-student design.

as in the previous academic year. We do not consider repeat matches to occur if a student has the same teacher in consecutive years, but not in the same subject. We also do not consider repeat matches to occur if a student returns to the same teacher after a gap.[6]

Students may repeat a grade due to academic failure. Grade retention depends on the students' performance during the school year, as well as their attendance rate. The most prevalent condition for grade retention between grades 4 and 8 is to fail (score below 4.0) in one subject and having a Grade Point Average (GPA) across all subjects lower than 4.45. Students must also attend at least 85% of classes. Grade retention is rare: about 1.8% of the students in grade 8 are repeating the grade. We do not exclude grade repeaters from our analysis because we implement a within-student comparison, as explained in Section 3.

Table 1 presents descriptive statistics. Panel (a) shows that the outcome (SIMCE test score) and treatment (repeated match) are measured at the student-subject level in grade 8. Repeat matches are common in the 8th grade of Chilean schools.[7] In the estimation sample, 58% of the observations have a repeat match. Panel (a) also shows that repeat matches are less common between grades 6 and 7 (41%) than between grades 7 and 8.[8]

There are no substantial differences in the frequency of repeat matches by subject, shown in panel (b). Because each student has probability of a repeat match of 0.58 in each subject, 8th graders can expect to have a repeat teacher in 2.32 of their four subjects. For each student we also observe sex, family background, past GPA, past attendance rate and class size in grade 8.

In panel (c) we report information at the teacher level, which includes sex, age and experience. Teachers' experience and age correspond to the average across the four

[6]Both are infrequent cases. In the sample, 88.9% of the total matches occur in the same subject. On the other hand, 2.8% of the student-teacher matches in 8th grade present 1 year of gap.

[7]Grade 8 is the final year of primary education, and students will typically move to a different school and have different teachers in grade 9. Students typically remain in the same school between grades 5 and 8, and therefore repeated student-teacher interactions will be common in grades 6, 7 and 8. Our analysis focuses on grade 8 because of the availability of the SIMCE test score information.

[8]We cannot identify repeat matches between grades 5 and 6 for the entire sample because we do not have enrolment data for 2001.

Table 1. Descriptive statistics

	Mean	*Standard deviation*
(a) Student-subject level i,s (N= 2,785,928)		
SIMCE test score	0.00	1.00
1=Repeat match grade 8	0.58	0.49
1=Repeat match grade 6-7	0.41	0.49
(b) Student level i (N= 696,482)		
1=Repeat match (Spanish)	0.57	0.50
1=Repeat match (Mathematics)	0.59	0.49
1=Repeat match (Natural Sciences)	0.59	0.49
1=Repeat match (Social Sciences)	0.58	0.49
Number of repeat matches	2.32	1.30
1=Female	0.51	0.50
Mother's schooling (years)	10.95	3.75
Household's monthly income (000s of CLP)	376.02	468.90
Past GPA	0.09	0.95
Past attendance rate (%)	94.40	5.81
Class size	26.68	8.47
(c) Teacher level j (N= 46,256)		
1=Female	0.68	0.47
Experience (average)	16.34	12.53
Age (average)	43.59	11.80
(d) School level k (N= 6,260)		
1=Public	0.50	0.50
1=Voucher	0.42	0.49
1=Private	0.07	0.26
1=SES 1 (Low)	0.25	0.43
1=SES 2 (Middle-low)	0.33	0.47
1=SES 3 (Middle)	0.23	0.42
1=SES 4 (Middle-high)	0.12	0.33
1=SES 5 (High)	0.07	0.25
1=Urban	0.73	0.44
School enrolment (average)	436.90	402.15
Number of teachers (average)	19.30	14.17
(e) Class-subject level c,s (N= 37,992)		
1=Problems to start the class	0.34	0.47
1=Classroom disruption	0.44	0.50
1=High teacher expectation	0.55	0.50

Notes: Sample comprises students in 8th grade in 2004, 2007, 2009 and 2011 who have valid test scores and a complete set of information on characteristics. Household monthly income is imputed from the mid-point of 15 income bands with widths of 100,000 CLP or 200,000 CLP. The class-subject information in panel (e) is only available for a subset of 9,498 classes out of 31,837 classes in total.

years.[9]

In panel (d) we report information at the school level including size according to enrolment and number of teachers. Schools in Chile may be one of three types: public, private but supported by vouchers and unsupported private.[10] Schools are classified by the Ministry of Education according to the socio-economic status (SES) of their students, based on four variables: father's level of education, mother's level of education, monthly family income and a vulnerability index of the students. The variable ranges between 1 and 5, 5 being indicative of the wealthiest students. Finally, in panel (e) we show information from the SIMCE survey about teachers' perceptions of classroom behaviour[11] and their expectations of their students in the future.[12]

In Table 2 we show how the characteristics of the treatment and control groups differ. The raw difference in test score is very small, but repeat matches are positively associated with several factors correlated with *worse* academic performance, including lower family income and lower previous test scores.

Panel (a) shows that repeat matches in grade 8 are themselves correlated with repeat matches in grade 7, which may reflect differences at the school-level in terms of policy towards repeated matches. However, the distribution of repeat matches does not suggest that looping is primarily a school-level policy. Two-thirds of students have variation in repeat matches across subjects (which by definition are taken within the same school). In Appendix A we show that only 15% of the variation in the proportion of repeat

[9]In the estimation sample teachers are observed a different number of times across the four years: 52% (24,271 teachers) are observed once; 24% (11,276 teachers) are observed twice; 14% (6,558 teachers) are observed three times, and and 9% (4,151 teachers) are observed four times.

[10]For a detailed description of the Chilean school system and education providers, see Santiago, Fiszbein, Jaramillo & Radinger (2017).

[11]Teachers were asked about how much they agree or disagree with the following statements: "In this class, it is very hard to start the class lessons" and "In this class, the lessons are often interrupted because I must silence or scold students". The rating scale is "I fully agree", "I agree"; "Disagree", "I entirely disagree". Both variables were coded as dummy variables, taking value of one if the teacher answers "I fully agree" or I agree" and zero otherwise.

[12]Teachers were asked "What do you think will be the highest level of education that most students in this class will achieve in the future?". The variable was coded as a dummy variable, taking value of one if the teacher expects that the majority of the class will complete higher education studies and zero otherwise.

matches at the school-subject-grade-year level is accounted for by school fixed effects, and also that very few schools always (or never) use repeat matches.

Panel (b) shows that students who have repeated matches come from lower-income families with less-educated mothers. Repeat matches are positively selected on those measures of academic effort and achievement which are observable by the teacher: past GPA and past attendance rate are both higher for repeat matches. However, repeat matches are *not* positively selected on the anonymised SIMCE test score.[13]

Panel (c) of Table 2 shows that repeat matches are significantly more common in public schools, in low socio-economic status schools and in rural schools. There are also important differences in terms of school size and structure, some of which are mechanically related to the probability of repeat matches. Students in smaller schools in terms of enrolment, number of classes, number of teachers and number of teachers per subject are all more likely to have repeat matches. Holding other factors constant, a reduction in the number of teachers who are available to teach a particular subject will increase the probability of repeat matches.

Panel (d) shows that repeat matches have significantly older and more experienced teachers. Repeat matches have teachers with three more years of experience than new matches in 7th grade (i.e. before the current match). Repeat matches have teachers with *six* more years of experience than new matches in 8th grade. More experienced teachers are more likely to get repeat matches, and, by definition, repeat matches have a teacher with one more year of experience than in the previous year. In contrast, new matches draw a new teacher who has more than two years *less* experience than their teacher in the previous year. This arises because, by definition, teachers who have repeat matches in 8th grade must have worked at the school in 7th grade, whereas new matches may draw a teacher who is new to the school.

[13]The SIMCE test is taken every year in 4th grade, from 2005 onwards. Therefore, past SIMCE test scores are only available in 2009 (4th grade in year 2005) and 2011 (4th grade in year 2007). 4th grade SIMCE scores are only available for three of the four subjects (Spanish, maths and natural sciences). As with current SIMCE test scores, scores in 4th grade are standardised to have mean zero and unit variance.

Table 2. Characteristics of treatment and control groups

	Treatment group (same teacher in grade 8)	Control group (new teacher in grade 8)	Difference	Std. err.
SIMCE test score	0.001	−0.002	0.003***	(0.001)
(a) Previous repeat matches				
1=Repeat match grade 6-7	0.47	0.32	0.154***	(0.001)
(b) Student characteristics				
1=Female	0.51	0.50	0.001	(0.001)
Mother's schooling (years)	10.74	11.26	−0.521***	(0.005)
Household's monthly income	342.37	422.66	−80.287***	(0.567)
Past GPA	0.11	0.06	0.050***	(0.001)
Past attendance rate (%)	94.62	94.09	0.536***	(0.007)
Past SIMCE test score	0.15	0.21	−0.058***	(0.002)
Class size	26.94	26.33	0.613***	(0.010)
(c) School characteristics				
1=Public	0.55	0.44	0.110***	(0.001)
1=Voucher	0.41	0.49	−0.079***	(0.001)
1=Private	0.05	0.08	−0.032***	(0.000)
1=SES 1 (Low)	0.11	0.09	0.027***	(0.000)
1=SES 2 (Middle-low)	0.34	0.30	0.044***	(0.001)
1=SES 3 (Middle)	0.35	0.35	−0.000	(0.001)
1=SES 4 (Middle-high)	0.15	0.19	−0.037***	(0.000)
1=SES 5 (High)	0.05	0.08	−0.033***	(0.000)
1=Urban	0.88	0.91	−0.035***	(0.000)
School enrolment	698.74	820.18	−121.432***	(0.741)
Number of classes	20.09	23.34	−3.248***	(0.018)
Number of teachers	26.29	31.01	−4.722***	(0.023)
Number of subject-teachers	2.66	3.20	−0.542***	(0.002)
(d) Teacher characteristics				
1=Female	0.69	0.68	0.011***	(0.001)
Experience in 7th grade	20.06	16.93	3.124***	(0.015)
Experience in 8th grade	21.06	15.37	5.694***	(0.014)
Δ Experience	1.00	−1.57	2.570***	(0.012)
Age	47.51	42.55	4.962***	(0.013)
Observations	1,618,387	1,167,541		

Notes: The past SIMCE test score is the SIMCE score from grade 4, and is based on 338,941 and 440,192 observations in the control and treatment groups respectively. All comparisons are at the student-subject level. The number of subject-teachers is based on the number of teachers in the school between 5th grade and 8th grade, because the majority of the teachers from the first cycle (grades 1–4) are general teachers, and they teach all the main subjects to a particular class. In the case of the four years analysed (2004, 2007, 2009, 2011), 95% of the teachers from the first cycle teach more than one subject. In contrast, 44% of the teachers from 5th grade to 8th grade are subject specialist, and teach only one subject. $^{*}p < 0.10$, $^{**}p < 0.05$, $^{***}p < 0.01$.

Given these differences in students, schools and teachers between repeat matches and new matches, it is important to note that we observe the same student (by definition in the same school) in multiple subjects, some of which are repeat matches and some of which are new matches, and we observe the same teacher with multiple classes,[14] some of which are repeat matches and some of which are new matches. This enables us to control both for unobserved fixed student effects and unobserved fixed teacher effects, which greatly reduces any concerns about selection on the basis of these characteristics.

3 The effect of repeat matches at the student-subject level

As shown in Table 2, a simple comparison of repeat matches and new matches may be misleading because repeat matches are not randomly assigned: repeat matches have systematically different students, teachers and schools. These differences may arise because of teacher and student sorting within schools, and because of teacher and student mobility between schools. Previous research has established the existence of teacher sorting within schools: less-experienced, minority and female teachers are systematically sorted to lower-performing students (Clotfelter, Ladd & Vigdor 2005, 2006, Feng 2010, Kalogrides, Loeb & Béteille 2013). Moreover, qualitative research shows that school leaders base their staffing decisions on a combination of teachers' performance (measured by their students' test scores) and teachers' preferences (Cohen-Vogel 2011, Kalogrides et al. 2013, Osborne-Lampkin & Cohen-Vogel 2014). Teacher and student mobility between schools may also cause differences in the proportion of repeat matches, and it seems likely that the decision to move schools will not be exogenous with respect to student outcomes.

Our data allow us to control for differences in fixed student characteristics by using

[14] A small fraction of teachers are observed in more than one school.

the within-student variation across subjects, taking advantage of the fact that we observe students' test scores in four different subjects.[15] In addition, since students attend the same school and the same class for all subjects, student fixed-effects will also control for selection bias as a result of differences in school or class characteristics. The inclusion of student fixed effects also addresses two specific sources of selection bias: parental choice of school and grade retention. First, parents' decision whether to move their child to another school could lead to a selection issue if parents take this decision based on, for instance, how well their children are matched with their teachers in a particular school. In the estimation sample 7.8% of the students change school between grade 7 and grade 8. Second, students who repeat the grade due to academic poor performance are significantly less likely to have a repeat match. In the estimation sample, about 1.8% of the students are grade repeaters, of which 65.7% do not have the same teacher again. Grade repeaters are more likely to come from low-income families, to have less educated mothers, and to have lower test scores. The inclusion of student fixed effects deal with both these potential biases, since children attend the same school for all subjects, and grade repeaters re-take all subjects.

As well as addressing selection bias, the inclusion of student fixed-effects allows us to estimate the effectiveness of repeat-matches independent of any effect of a group of students staying together between grades. It seems possible that student-student familiarity (in addition to student-teacher familiarity) has a causal effect on student outcomes, and the process of assigning the same teacher to a group of children necessarily implies that the group (or at least the majority of the group) stay together between grades. The fixed-effect strategy we use compares the same student across subjects in the same year, and this student will have the same classmates for all subjects, so we are effectively comparing outcomes for the same group of students, some of whom have a repeat match and some of whom do not.

[15] Many cross-sectional studies exploit within-student variation to identify effects of teacher characteristics and teaching practices (Dee 2007, Clotfelter, Ladd & Vigdor 2010, Bietenbeck 2014, Bietenbeck, Piopiunik & Wiederhold 2018, Paredes 2014, Lavy 2015, Comi, Argentin, Gui, Origo & Pagani 2017).

Our method also allow us to control for differences in fixed teacher characteristics by using the within-teacher variation across classes, taking advantage of the fact that we observe the same teacher in several classes. Further, and in contrast to students, we observe the same teacher in multiple classes at four different points in time (2004, 2007, 2009 and 2011) which allows for the inclusion of teacher-by-year fixed effects. As was clear from Table 2, there is inevitably a strong relationship between repeating the student-teacher match and teacher experience. Even if repeat-match teachers were drawn randomly, these teachers by definition must have worked in the same school at $t-1$, but new match teachers are drawn from the pool of available teachers which includes those who are new to the school. In addition, repeat-match teachers are not drawn randomly: they have about three more years of experience, on average. Thus, an unconditional comparison of classes which have a repeat match with those that do not conflates the advantages of a repeat match with any advantages of having a teacher who has nearly six years more experience (see panel (d) of Table 2). Since experience is fixed for a given teacher in a given year, the inclusion of teacher-by-year fixed effects controls for this large difference in experience.

Thus, our first model to identify the effect of a repeat match is:

$$y_{is} = \beta_1 R_{is} + \mu_i + \mu_s + \mu_{jt} + \varepsilon_{is}, \tag{1}$$

where y_{is} is the standardised SIMCE test score of student i in grade 8 in subject $s = 1,2,3,4$ (maths, Spanish, social sciences, natural sciences). Each student is observed in grade 8 in one year $t = 2004,2007,2009,2011$, and therefore i identifies t. For a particular student-subject-year combination we observe the identity $j = J(i,s,t)$ of the teacher. In (1) each student i appears in only one school in one year, whereas teachers j appear in multiple classes and years and may also be observed in more than one school. R_{is} is an indicator variable which takes the value 1 if there is a repeat match, which occurs if $J(i,s,t-1) = J(i,s,t)$. As discussed, the model includes student, subject and teacher-by-year fixed effects.[16]

Table 3 presents estimates of versions of Equation (1) with the inclusion of different fixed-effects. Across all specifications, the results show a positive and significant effect of repeating the student-teacher match on student's SIMCE test scores. The raw effect in Column (1) is small, but recall from Table 2 that repeated matches are far from randomly assigned, and are often associated with baseline characteristics which themselves are associated with lower test scores. Including student fixed effects in column (2) increases the effect to 0.026σ, while the inclusion of both student and teacher effects in columns (3) and (4) reduces the effect to 0.017σ. The inclusion of teacher-by-year fixed effects in column (4) controls for any effect of differential experience between teachers who repeat matches and those who do not and increases the estimate to 0.019σ.[17] We find no evidence that the size of the effect varies across subjects: an F-test of the interactions between R_{is} and μ_s is insignificantly different from zero. It is also possible to replace the teacher-by-year fixed effects μ_{jt} with teacher-by-subject-year fixed effects μ_{jst} to ensure that we are not conflating looping with an effect from non-looping teachers teaching different (possible less preferred) subjects. The inclusion of μ_{jst} slightly reduces the estimate to 0.016σ.

In column (5), we include as a control lagged test scores at the student-subject level (Rivkin et al. 2005, Harris & Sass 2011, Chetty, Friedman & Rockoff 2014). This is a value-added model which controls for within-student differences in ability across subject which may be correlated with the looping decision. However, the SIMCE test score information for these students is only available in grade 4 and grade 8, so this does not deal with the problem that the decision to loop may be based on match quality in grade 7. The sample in column (5) is significantly smaller because the grade 4 SIMCE score is only available in 2009 and 2011, and only in three of the four subjects.[18] The

[16] The model is estimated using the methods developed by Correia (2016) and Guimaraes & Portugal (2010).

[17] Excluding students who have no variation in R across subjects makes almost no difference, with an estimated effect of 0.018 (0.002).

[18] Repeating the column (4) model on this reduced sample yields an estimate of 0.020 (0.004).

Table 3. Effect of repeat student-teacher match on test scores: fixed-effect estimates

	(1)	*(2)*	*(3)*	*(4)*	*(5)*	*(6)*
Repeat match grade 7-8	0.003**	0.026***	0.017***	0.019***	0.021***	0.020***
$R_{is} = 1$	(0.002)	(0.001)	(0.001)	(0.002)	(0.004)	(0.004)
SIMCE score in grade 4					0.276***	0.276***
					(0.002)	(0.002)
$R_{is} = 1$ grade 6-7						0.014***
						(0.003)
$R_{is} = 1$ grade 5-6						0.007***
						(0.002)
$R_{is} = 1$ grade 4-5						0.006
						(0.004)
Subject FE	Yes	Yes	Yes	Yes	Yes	Yes
Student FE		Yes	Yes	Yes	Yes	Yes
Teacher FE			Yes			
Teacher FE × Year FE				Yes	Yes	Yes
R-squared	0.000	0.793	0.808	0.812	0.849	0.849
Observations	2,785,928	2,785,928	2,785,928	2,785,928	759,597	759,597

Notes: Dependent variable is the student's SIMCE test score in grade 8. In all columns, treatment is the student-subject measure of repeated match R_{is} in grade 8. Standard errors are clustered at the student level. $^{*}p < 0.10$, $^{**}p < 0.05$, $^{***}p < 0.01$.

inclusion of lagged SIMCE scores makes almost no difference to the estimate. Finally in column (6) we deal with the concern that repeat matches may be correlated with earlier looping decisions by including as controls the value of R_{is} in grades 5, 6 and 7. Once again, this makes almost no difference to our estimate of the effect of looping on test scores in grade 8.

Our estimates are very similar to those reported by Hill & Jones (2018) for younger students' maths scores in North Carolina elementary schools (grades 3–5) using a similar specification, but which also includes lagged test scores as a control variable.[19] The outcome measure used by Hill & Jones is a maths score which was reported by the teacher themselves, rather than an anonymised national test score as in our case. This suggests that the use of an anonymised test score, as in our case, is not crucial for finding

[19]Hill & Jones (Table 2) report an effect size of 0.018σ (0.005). The increased precision of our estimates likely reflects the much wider prevalence of repeat matches in our data; Hill & Jones report that only three percent of students experience a repeat match in their data.

positive effects from repeated matches.

The remaining source of variation in (1) is the error term ε_{is}, which varies at the student-subject (equivalent to the student-teacher) level. If repeat matches are formed non-randomly with respect to this "match quality" term, then estimates of β_1 will still be biased even after controlling for student and teacher fixed-effects. Schools or parents may both make decisions about which class-teacher matches to keep together in grade 8 on the basis of their performance in grade 7. As a result, class-teacher matches are endogenously destroyed and the effect of a repeat match will be confounded by survivor bias.

Unfortunately, we do not have the SIMCE test score in grade 7 for these students. However, we can use information on SIMCE scores in grade 6 to predict match formation in grade 7. To do this we estimate (1) on a sample of all grade 6 students for whom we have SIMCE test scores[20] and calculate $\hat{\varepsilon}_{is,6}$, the residual for each student-subject observation. We then calculate, for each student-subject observation, the average residual of their classmates, $\overline{\hat{\varepsilon}}_{i's,6}$ and estimate whether these residuals have any effect on the formation of repeat matches in grade 7:

$$R_{is,7} = \gamma_1 \hat{\varepsilon}_{is,6} + \gamma_2 \overline{\hat{\varepsilon}}_{i's,6} + \mu_{jt} + \eta_{is,7}. \tag{2}$$

In this model, γ_1 captures whether students whose individual residual is high are more likely to remain with the same teacher in grade 7, while γ_2 captures whether students whose classmates have high residuals are more likely to remain with the same teacher in grade 7. Our estimate of γ_1 is negative, but extremely small and insignificantly different from zero (-0.0003 (0.0004)). Our estimate of γ_2 is slightly larger but still insignificantly different from zero at conventional levels (-0.009 (0.005)). Thus, we find no evidence that student-subject combinations which perform better than expected are more likely to lead to repeat matches.

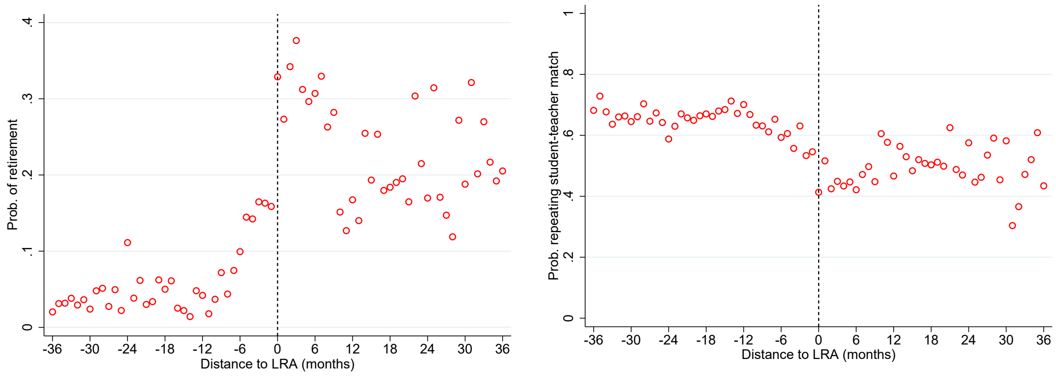
[20]We have information on SIMCE scores for Spanish and Math in 5 years (2013, 2014, 2015, 2016 and 2 > 018).

Nevertheless, because we cannot directly control for endogenous selection, we also consider a regression discontinuity approach which exploits the discontinuity in the probability of a repeat match which occurs because of small differences in teachers' date of birth in the year before the grade 8 observation which affect exactly when teachers reach the legal retirement age (LRA). A student whose teacher reaches the LRA in grade 7 is less likely to match in grade 8, because that teacher is more likely to retire. The discontinuity which occurs at the LRA is plausibly exogenous with respect to ε_{is}. Clearly, the retirement decision itself is unlikely to be exogenous with respect to student performance, as noted by Fitzpatrick & Lovenheim (2014). Hanushek, Kain & Rivkin (2004) also argue that there are teacher selection effects with age which can bias estimates of the returns to teacher experience. However, although match (or teacher) quality may vary with teacher age, there is no reason why they would be discontinuous at the LRA itself. Manipulation of (reported) teacher date of birth is implausible in this setting.

In Chile, the LRA is 65 for men and 60 for women, but teachers are not obliged to retire from the labour market at that age. The law permits early retirement, provided that teachers meet some financial requirements.[21] The school-year starts during the first week of March and finishes in late November or early December. School administrators assign teachers to classes on the assumption that teachers will remain in the school until the end of the school year. Each teacher's exact date of birth is recorded, and using this we calculate age for each teacher on the last day in February in each year (2004, 2007, 2009 and 2011), i.e. the day before the school year starts. Our key identifying claim is that teachers who reach the LRA just before the 1 March are significantly more likely to retire than teachers who reach the LRA just after 1 March. For example, a grade 7 class in the 2006 school year whose (female) teacher reaches 60 in February 2007 is less likely to have the same teacher in grade 8 than a class whose teacher reaches 60 in

[21] To retire early, workers are required to have sufficient pension resources to fund a replacement rate of 70 percent with respect to their average salary over the previous 10 years, and a minimum pension set by law.

Figure 1. Discontinuity in retirement at the LRA and repeat matches, distance in months



(a) Probability of retirement

(b) Probability of repeating the student-teacher match

Notes: A teacher is considered retired if she does not appear in the next five consecutive years in the administrative records of Ministry of Education. The distance to the legal retirement is the difference between the current age and the LRA, recorded in months. The distance to the legal retirement is zero for those teachers whose birthdays are in February and therefore reach the LRA in the last month of the previous school year.

March 2007.

Although we do not have a formal measure of retirement, we observe the population of school-teachers in Chile in each year and therefore we can infer retirement quite precisely from the disappearance of a teacher from the data for the next five years. In the left-hand panel of Figure 1 we show that the the probability of retirement increases quite sharply (but with no discontinuity) for teachers who will reach the LRA in the next school year, and then jumps by over 10 percentage points between teachers who reach the LRA in February (distance to LRA$=0$) and those who reached it in March (distance to LRA$=-1$). In the right-hand panel of Figure 1 we show that this discontinuity is reflected in a sharp 15 percentage point reduction in the probability of a repeat match.

We therefore have an fuzzy-RD design with distance to the LRA of each student-subject combination in grade 7, denoted D_{is}, as the running variable, which can be measured in days. Following Imbens & Lemieux (2008) the RD estimator is defined as:

$$\tau_{RD} = \frac{\lim_{D_{is}\downarrow 0} E[y_{is}|D_{is}=0] - \lim_{D_{is}\uparrow 0} E[y_{is}|D_{is}=0]}{lim_{D_{is}\downarrow 0} E[R_{is}|D_{is}=0] - \lim_{D_{is}\uparrow 0} E[R_{is}|D_{is}=0]} = \frac{\tau_y}{\tau_R} \tag{3}$$

As before, y_{is} denotes the SIMCE test score in 8th grade. The RD estimator corresponds to the ratio between the average intention-to-treat effect (τ_y) and the first-stage effect (τ_R).

We adopt a local polynomial modelling approach to approximate the functional form of τ_y and τ_R. This method uses only the observations that lie between $-h$ and $+h$, where h is a positive bandwidth. Local polynomial estimation involves choosing a kernel function to weight the observation within the the interval $[-h, +h]$. We use a triangular kernel function, which gives the maximum weight at $D_{is} = 0$. We use a polynomial of order one, that is to say, we run a local-linear regression within the bandwidth. To select the bandwidth we follow the procedure proposed by Calonico, Cattaneo & Titiunik (2014) by selecting the parameter h that minimises an approximation to the asymptotic mean squared error (MSE) of the point estimator ($\hat{\tau}^{RD}$). Intuitively, choosing a small bandwidth will reduce the approximation bias, but at the same time will increase the variance of the estimated coefficient. For inference, we use robust confidence intervals based on bias-correction following Calonico et al. (2014).

The validity of the discontinuity approach is based on the usual three IV assumptions. First, a relevance condition, that the LRA has a strong effect on the probability of teacher retirement, which in turn affects the probability of repeating the student-teacher match. We have already seen that the discontinuity is a powerful predictor of retirement, and therefore of repeat matches. Second, the instrument exogeneity condition, in this case that the discontinuity at the LRA is exogenous with respect to student potential outcomes. In Figure B1 in Appendix B we provide evidence that differences in observable characteristics either side of the LRA are very small and almost all insignificantly different from zero compared to the differences in the treated and controls. Figure B2 shows that density of the running variable shows no sign of manipulation at the cut-off.[22] In order to deal with any remaining imbalance we supplement our RD estimates

[22] The manipulation test of Cattaneo, Jansson & Ma (2020) estimates the density of the running variable either side of the cutoff using a local polynomial and yields a p-value of 0.1314.

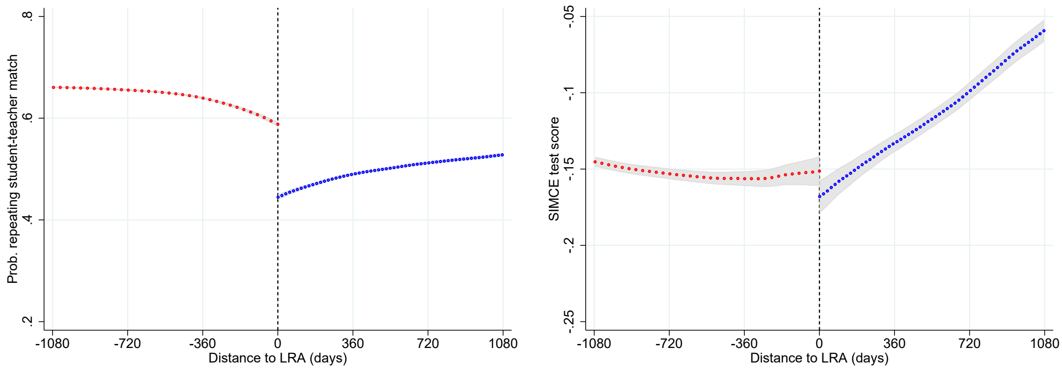
with parametric RD estimates which allow for within-student and within-teacher comparisons. Third, we require that the discontinuity effect on student outcomes is *only* driven by its effect on repeat matches. There are two threats to the exclusion restriction. Even if the variation in repeat matches which is caused by the discontinuity is as good as randomly assigned, this variation also causes (quite large) variation in teacher experience. To deal with this, we also consider parametric RD models which allow for the inclusion of teacher-by-year fixed effects which remove any variation in experience between repeated and non-repeated classes.

The resulting RD estimates are local for a very specific type of repeat match. The discontinuity will identify the causal effect of a repeat match with an experienced teacher who complies with the discontinuity. In other words, a teacher whose retires at the LRA. If the effect of repeat matches itself varies with teacher experience, then the IV estimates will not be comparable to the fixed-effect estimates from (1).

The regression discontinuity results are illustrated in Figure 2, which shows the first stage estimate of τ_R in the left-hand panel and the reduced form estimate of τ_y in the right-hand panel. As we anticipated, the first stage shows a large negative effect: students whose teacher reaches the LRA in grade 7 are about 17 percentage points less likely to repeat the match in grade 8. The reduced-form effect on SIMCE test score is about -0.03σ: students whose teacher reaches the LRA in grade 7 have lower test score outcomes in grade 8.

In Appendix B (Figure B1) we provide some evidence on the exogeneity assumption by estimating the non-parametric RD model but using a wide range of measured characteristics as the outcome variable. For reference, we also show the estimated difference in means from a raw comparison of treated and controls. In the top panel, differences in means are greatly reduced and in most cases insignificantly different from zero. However, some small imbalance remains. One possible explanation for this is that early retirement decisions may also be discontinuous at the February-March threshold, and

Figure 2. Conditional mean plots by local linear regressions: probability of repeating student-teacher match and SIMCE test score



Notes: Panel a) the probability of repeating the student-teacher match against the distance to the LRA between $[-1{,}080, 1{,}080]$ days. Panel b) SIMCE test scores against the distance to the LRA between $[-1{,}080, 1{,}080]$ days. The distance to the legal retirement is the difference between the current age and the LRA. The distance to the legal retirement is zero for those teachers whose birthdays are 1st March and reach the LRA in that day. The graphs show conditional mean plots using local linear regression within a MSE-optimal bandwidth (bandwidth = 965 days), with triangle kernel function and a 1st order polynomial, on a grid of 500 points on each side of the cutoff.

those decisions may be related to school type.[23] In the bottom panel we repeat the exercise but include controls for school type (public, private, voucher). We now see even less imbalance across the discontinuity. The only exception remaining is household income, which is slightly higher for children whose teacher's age is just below the LRA. As noted, income in the SIMCE data is reported in 15 bands from which we imputed a continuous variable. All of these bands are balanced across the discontinuity once we control for school type, as shown in Table B1. The possibility that there are small imbalances at the discontinuity motivates us to also consider parametric RD models which allow for within-school and within-teacher comparisons.

Figure 2 implies a causal effect of repeat matches which is substantially larger than the fixed-effect estimates in Table 3, because the ratio of τ_y and τ_R is approximately 0.2σ. In Column (1) of Table 4 we report a non-parametric RD estimate of 0.158σ which corresponds exactly to Figure 2. However, this large estimate may arise because

[23]For example, if some schools encourage teachers to retire at the end of the school year *before* they reach the LRA, there may be imbalance in characteristics at that threshold in the following year.

we are conflating the repeat-match effect with an experience effect: although the discontinuity as good as randomly selects students into repeat matches, the discontinuity also selects students into more or less experienced teachers. We test whether this large estimate is due to the experience effect by applying exactly the same RD model to teacher experience. Our estimate of the teacher experience effect of the discontinuity is very large: over 21 years with a standard error of less than one year. This means that, although we can plausibly claim that the LRA discontinuity as good as randomly breaks up student-teacher pairs in grade 7, it has a large causal effect both on the probability of repeating the match *and* on the experience of the teacher in grade 8.

Therefore, in column (2) of Table 4 we we adopt a linear functional form for the distance to the LRA, which has a number of advantages. First, it greatly improves estimation precision. Second, and more importantly, it allows us to include student and teacher-by-year fixed-effects, which sweep out any non-random selection of new teachers in comparison to the teachers of continuing matches. In particular, it allows us to control for the experience effect of looping. As expected, this method reduces the effect of looping and produces an estimate of 0.110σ with a substantially smaller standard error. Our estimate of the returns to experience suggests that about half the difference between the results in column (1) and (2) can be accounted for by the loss of experience which is associated with getting a new teacher in grade 8.[24]

A disadvantage of the simple linear model reported in column (2) is that Figure 2 suggests that the relationship between looping and age is somewhat non-linear in the year before and after the LRA. Therefore, in column (3) we report a quadratic model which allows for this non-linearity but which also allows for the inclusion of student and teacher by year fixed-effects. The quadratic model yields an estimate of 0.124σ, with a slightly larger standard error than the linear model.

[24]Our data allows us to estimate the likely effect of this loss of experience since we have a clean measure of student achievement and teacher experience. Following the method of Harris & Sass (2011) our return to experience model in Appendix C predicts that losing a teacher at the LRA with 25 years experience (the sample mean) and replacing them with a new teacher causes a loss in student test scores of 0.024σ.

Table 4. Effect of repeating the student-teacher match on test scores: regression discontinuity results

	Non-parametric *(1)*	*Linear with fixed effects* *(2)*	*Quadratic with fixed effects* *(3)*
τ_R (First stage)	-0.137^{***}	-0.121^{***}	-0.114^{***}
	(0.004)	(0.004)	(0.005)
τ_y (Reduced form)	-0.022^{**}	-0.013^{***}	-0.014^{**}
	(0.009)	(0.005)	(0.006)
τ_{RD}	0.158^{**}	0.110^{***}	0.124^{**}
	(0.063)	(0.038)	(0.051)
Student FE		Yes	Yes
Subject FE		Yes	Yes
Teacher FE $\times$ Year FE		Yes	Yes
First-stage R-squared		0.873	0.873
First-stage F statistic		1,041	566
95% C.I.	[.035 ; .28]		
Effective observations: Left	200,343		
Effective observations: Right	109,731		
Optimal Bandwidth	964.830		
Observations	2,785,928	2,785,928	2,785,928

Notes: Dependent variable is the student's SIMCE test score in grade 8. Treatment is the student-subject measure of repeated match R_{is}. Column (1) presents results based on Calonico et al. (2014) with a polynomial of order one and weighted by a triangular kernel. Column (2) includes distance to the LRA linearly, and the interaction between the distance to the LRA and the indicator variable for reaching the LRA. Column (3) includes a quadratic interaction between distance to the LRA linearly and the indicator variable for reaching the LRA. Standard errors in Column (1) are calculated using Calonico et al. (2014). Standard errors are clustered at the student-level. $^{*}p < 0.10$, $^{**}p < 0.05$, $^{***}p < 0.01$.

Do the positive effects of repeated matches occur for every subject? In Appendix D we investigate this issue by estimating the linear RD model separately for Spanish, maths, natural sciences and social sciences. In these models we cannot control for student fixed effects because each student is observed only once in each subject in grade 8, but we can still control for teacher-by-year fixed effects because teachers take multiple classes in the same subject (both within and across years).

In all four subjects there is a strong negative effect of reaching the LRA on the probability of repeating the match. This effect is weaker in Spanish, but very consistent in the other three subjects. The reduced form estimate of τ_y is negative in all four subjects, implying that the estimate of τ_{RD} is positive in all four subjects. However, standard errors are considerably larger than in the equivalent linear model because the sample size is much smaller, so it is hard to make precise statements about the difference in effectiveness across subjects. The effect appears smallest in natural sciences and largest in Spanish, but these results are too imprecise to draw more conclusions about the efficacy of repeat matches in different subjects.

All our RD estimates are larger than the fixed-effects and value-added estimates. This seems unlikely to be the result of strong negative selection into repeat matches. The RD estimates are local in that they relate to very experienced teachers whose retirement decision is affected by reaching the LRA. Therefore, our results suggest that the benefits of looping may be significantly greater for more experienced teachers. However, a natural concern is that, instead, this reflects a failure of the exclusion restriction. Since our parametric models include grade 8 teacher-by-year fixed effects, any failure of the exclusion restriction can only plausibly come from discontinuities in grade 7. A particular concern is that the discontinuity may have an effect on teacher effort in grade 7 which may in turn effect outcomes in grade 8. We test of this restriction by considering a sample of students who change school between grade 7 and grade 8. These students cannot loop,[25] and their grade 8 teacher is selected independently of the grade

[25] A tiny number of school movers do in fact have the same teacher in grade 8, presumably because

Table 5. Effect of the discontinuity on school-movers: reduced form regression discontinuity results

	Non-parametric (1)	*Linear with fixed effects (2)*	*Quadratic with fixed effects (3)*
τ_y	-0.022	-0.0004	-0.010
	(0.029)	(0.010)	(0.013)
Student FE		Yes	Yes
Subject FE		Yes	Yes
Teacher FE $\times$ Year FE		Yes	Yes
95% C.I.	[-0.079;0.035]		
Effective observations: Left	181,754		
Effective observations: Right	11,706		
Optimal Bandwidth	1280.411		
Observations	193,460	193,460	193,460

Notes: Sample restricted to students who changed school between grade 7 and grade 8. Dependent variable is the student's SIMCE test score in grade 8. Column (1) presents results based on Calonico et al. (2014) with a polynomial of order one and weighted by a triangular kernel. Column (2) includes distance to the LRA linearly, and the interaction between the distance to the LRA and the indicator variable for reaching the LRA. Column (3) includes a quadratic interaction between distance to the LRA linearly and the indicator variable for reaching the LRA. Standard errors in Column (1) are calculated using Calonico et al. (2014). Standard errors are clustered at the student-level. $^{*}p < 0.10$, $^{**}p < 0.05$, $^{***}p < 0.01$.

7 discontinuity, which leaves grade 7 teacher effort as the only channel by which the discontinuity can effect test scores in grade 8.

Table 5 reports estimates of the reduced form τ_y when the sample is restricted to school-movers. All three estimates are insignificantly different from zero, although we note that the non-parametric estimate in column (1) are imprecise and of the same size as in Table 4. More encouragingly, the parametric estimates of τ_y are close to zero. These estimates support our claim that the effect of the LRA discontinuity on test scores operates through its effect on repeat matches.

their teacher moved simultaneously or because their teacher had classes in multiple schools.

4 The effect of repeated matches on students, classes and schools

The comparison we made in Section 3 was between individual student-teacher matches that repeat and those that do not. The great advantage of this comparison is that allows us to make within-student and within-teacher comparisons, and our RD strategy also allows to control for endogenous matches at the student-subject level. However, repeat matches may have spillover effects on untreated units. At the student level, a student may allocate greater effort to subjects in which there is a repeat match, but at the same time allocate less effort to non repeat-match subjects. If this was the case, increasing the number of matches at the student level would be less effective. At the class level, if repeat matches allow teachers to save time, there will be benefits to all students in the class, regardless of whether students are individually repeating the match. On the other hand, if repeat matches are beneficial because of greater familiarity between teacher and student, it might not be beneficial for those who join a class in which most other students have a familiar teacher. Indeed, it seems possible that it might actually be harmful if teachers focus their efforts on students with whom they are familiar. At school-level, the allocation of teachers is a joint problem where repeating a match for one teacher has some implication for all other allocations within that school. In this section we therefore aggregate our data and use fixed-effect methods to examine whether the positive effects at the student-subject level carry over to student, class and school-level.

Our student-level model is:

$$\bar{y}_i = \beta_1 \bar{R}_i + \beta_2 x_i + \mu_c + \varepsilon_i, \tag{4}$$

where $\bar{y}_i$ is student i's average SIMCE score across all four of their grade 8 subjects, and $\bar{R}_i$ is the proportion of their four subjects in which they have the same teacher as in grade 7. The model includes class fixed-effects μ_c and a set of pre-determined student-level characteristics x_i. The variation we are exploiting here is the within-class variation in repeat matches which arises because not all students in a particular class in grade 8 will have had the same teacher in grade 7.

Our class-subject model is:

$$\bar{y}_{cs} = \beta_1 \bar{R}_{cs} + \beta_2 x_j + \mu_c + \mu_s + \varepsilon_{cs}, \tag{5}$$

where $\bar{y}_{cs}$ is the average SIMCE score of all students in class c and subject s in grade 8, and $\bar{R}_{cs}$ is the proportion of the class-subject combination who have the same teacher as in grade 7. The model includes class μ_c and subject μ_s fixed-effects and a set of pre-determined teacher-level characteristics x_j. The variation we are exploiting here comes from that the fact that $\bar{R}_{cs}$ varies across subject within class. Note that in both (4) and (5) there is no time variation because each student and class is observed in only one year.

Finally, our school-subject-level model is:

$$\bar{y}_{kst} = \beta_1 \bar{R}_{kst} + \beta_2 x_{ks} + \mu_{kt} + \mu_s + \varepsilon_{ks} \tag{6}$$

where $\bar{y}_{kst}$ and $\bar{R}_{kst}$ are the school-subject-year level averages of y_{is} and R_{is} in Equation (1); μ_{kt} is a school-by-year fixed effect; μ_s is a subject fixed effect; x_{ks} is a vector of characteristics of the school that vary across subjects and years (specifically, the proportion of female teachers and average experience). The parameter of interest is β_1. Note that at the school level we have four cohorts of grade 8 students from 2004, 2007, 2009 and 2011, and hence (6) has time variation. Equation (6) relies on variation within schools across subjects and across time for identification. This allows us to rule out selection into schools which might occur if, for example, better schools have more (or less) repeat matches. Also, exploiting the fact that we observe the same school for different cohorts, it is possible to include a school-by-year fixed effect μ_{kt}. This effect will remove all differences between school cohorts which might arise if repeat matches are

Table 6. Effect of repeat matches on test scores at student, class and school-level

	Student level $\bar{R}_i$ *(1)*	*Class-subject level* $\bar{R}_{cs}$ *(2)*	*School-subject-year level* $\bar{R}_{kst}$ *(3)*
Proportion of repeat matches	0.039***	0.029***	0.032***
	(0.006)	(0.002)	(0.002)
Class FE	Yes	Yes	
Subject FE		Yes	Yes
School-by-year FE			Yes
Student controls	Yes		
Teacher controls		Yes	Yes
R-squared	0.414	0.916	0.911
Observations	696,482	127,348	82,524

Notes: In each model the dependent variable is the proportion of repeat matches at that level. Model (1) includes controls for students' gender, household income, mother's education and attendance rate in grade 7. Models (2) and (3) include controls for teachers' gender and experience. Standard errors are clustered at the class-level. $^{*}p < 0.10$, $^{**}p < 0.05$, $^{***}p < 0.01$.

used for some cohorts and are related to cohort-specific unobservable shocks.

All three estimates are positive and significant, consistent with a positive effect of repeat matches on students, classes and schools. It is striking that all three estimates are larger than the comparable student-subject level estimates in Table 3. This can partly be explained by the fact that these models do not control for teacher fixed-effects — the exclusion of teacher effects in Column (2) of Table 3 produces larger estimated effects at the student-subject level as well. Larger effects are also consistent with positive spillovers from repeat matches within students, classes and schools.

5 Classroom behaviour and teacher expectations

Our results consistently show that repeating the student-teacher match results in a positive effect on student test scores. We find these effects at various different levels of aggregation. In this section, we provide further evidence of the effectiveness of repeat matches on the behaviour of students and the views of their teachers. Specifically, we estimate the effect of repeat matches on student attendance, student behaviour and teacher expectations of their students.

The student enrolment data contains a record of student attendance measured at the student level (we not observe attendance by subject separately for each student), so we estimate a variant of (4) and regress the standardised attendance rate on $\bar{R}_i$, the proportion of subjects in which the student has a repeat match in grade 8. As in (4), the model includes class fixed effects and therefore relies on within-class variation.

An independent measure of student behaviour is available from the survey of teachers about their perception of classroom behaviour and the future performance of the class, which is available in 2009 and 2011. Although teachers who complete these surveys are clearly aware of whether their class is a repeat match or not, it is nevertheless a measure which is entirely independent of the anonymised SIMCE test score. Teachers do not know what their students' test scores are, and so this cannot influence their responses to the survey.[26] There are three survey responses of interest. Teachers are asked if they face behavioural problems at the beginning of the class and disruptions during the class. These two outcomes are coded as binary variables, taking value of 1 if they are strongly agree or somewhat agree, and 0 otherwise. In addition, teachers are asked about the level of education that most of the class will achieve. The teacher expectation is coded as a binary variable, taking value of 1 if the teacher expects the majority of the class would finish any type of higher education (either a professional degree or a technical degree) or postgraduate studies. Our data is at the class-subject level, so we use a variant of (5) where the dependent variable is our measure of teacher perception (behaviour, expectations) for class c subject s, and the treatment is $\bar{R}_{cs}$, the proportion of the class c that repeat the match in the subject s. Fixed effects at class level are included

[26] 41% of the classes in data have this survey information for each subject. Table E1 (Appendix E) reports a mean comparison test of classroom observable characteristics for the estimation sample and the restricted sample. The restricted sample has more socio-economically advantaged students, and also has students with a better average performance in the SIMCE test. Although the differences between the two samples are statistically significant, they are not large.

Table 7. Effect of repeat matches on student behaviour and teacher expectations

	Attendance	Problems to start the class	Classroom disruption	High teacher expectations
	(1)	*(2)*	*(3)*	*(4)*
Proportion of repeat matches	0.052***	−0.041***	−0.044***	0.017**
	(0.008)	(0.007)	(0.008)	(0.007)
Class FE	Yes	Yes	Yes	Yes
Subject FE		Yes	Yes	Yes
Student controls	Yes			
Teacher controls		Yes	Yes	Yes
R-squared	0.516	0.418	0.439	0.566
Observations	696,482	37,992	37,992	37,992

Notes: Model (1) is at the student level and include controls for students' gender, household income, mother's education and attendance rate in grade 7. Models (2), (3) and (4) are at the class-subject level and include controls for the teacher's gender and experience. Standard errors are clustered at the class level. $^{*}p < 0.10$, $^{**}p < 0.05$, $^{***}p < 0.01$.

to capture all the subject-invariant characteristics (observable and unobservable) of the class.

Results are displayed in Table 7. Column (1) indicates that repeat matches have a positive effect on attendance, increasing it by 0.05σ, an effect size which seems plausibly consistent with the effect on test scores. Repeat matches also improve the teacher's perception of classroom behaviour and teacher expectations, shown in Columns (2)–(4). In particular, teachers are 4.1 percentage points less likely to have behavioural problems at the beginning of the class and 4.4 percentage points less likely to experience disruptive student behaviour. There are smaller but still significant effects on teacher expectations: teachers are 1.7 percentage points more likely to hold higher expectations for their students if their class is entirely made up of repeated matches.

These results are consistent with the qualitative evidence from teachers who claim that "looping" is beneficial for classroom behaviour. Students are familiar with the expectations of behaviour set by the teacher in previous years, and as a result behaviour improves. Of course, we cannot tell if the positive effects of repeat matches are jointly responsible for improved student behaviour and improved test scores, or whether improved behaviour is a mechanism by which academic performance improves.

6 Conclusions

There is a large literature which stresses the importance of teacher quality for student outcomes. But teacher quality is hard to improve. In this paper, we have provided evidence that there are significant benefits to reallocating existing teachers to students they have taught before. Qualitative evidence from teachers suggests that repeating the match saves time, engenders greater familiarity, and hence aids learning. However, estimating the causal effect of student-teacher familiarity is challenging for two reasons. First, because student-teacher matches are non-randomly selected. Second, because, even if student-teacher matches were chosen randomly, a repeat match may affect student performance for reasons other than student-teacher familiarity: we have seen that repeat matches have more experienced teachers and may also have more within-class familiarity.

We have provided a range of evidence from a new setting to suggest that repeating the student-teacher match has a significant positive effect on student test scores: we consider older (grade 8) children in a situation where repeat matches are common. A multidimensional fixed-effects framework which controls for selection by student or teacher into repeat matches suggests that repeat matches have test scores about 0.02σ higher, a result which is very consistent with evidence for younger children from the US. Our results also support a wide range of case-study and qualitative findings from the educational literature. The fixed-effects methods effectively hold constant many of the other channels by which repeat matches might affect student outcomes. A regression discontinuity design which additionally controls for selection on the basis of subject-specific match quality suggests larger effects in the range 0.11σ to 0.16σ, albeit with much less precision.

We have also shown that these effects aggregate to the class and school-level, which implies that the positive effects for treated classes are not simply at the expense of untreated classes, which would be the case if, for example, schools simply allocate more effective teachers to repeat matches. Our final piece of evidence suggests that the effects continue over time, and that university test scores increase with the number of repeated matches over a student's school career. Consistent with our findings of positive effects on test scores, we also find positive effects in teachers' perceptions of classroom behaviour and their expectations of their students' achievements.[27]

Allocating teachers to groups of students with whom they have interacted in the past appears to bring significant improvements in student performance without incurring additional costs on schools. An important question for future research is whether these results, which are estimated from variation in repeat matches in observational data, can be verified in a randomised setting.

References

Aucejo, E. M., Coate, P., Fruehwirth, J. C., Kelly, S. & Mozenter, Z. (2018), Teacher effectiveness and classroom composition. CEP Discussion Paper 1574, Centre for Economic Performance, London School of Economics and Political Science.

Bietenbeck, J. (2014), 'Teaching practices and cognitive skills', *Labour Economics* **30**, 143–153.

Bietenbeck, J., Piopiunik, M. & Wiederhold, S. (2018), 'Africa's skill tragedy: Does teachers' lack of knowledge lead to low student performance?', *Journal of Human Resources* **53**(3), 553–578.

[27]Note that our measure of test scores comes from an anonymous national test which is not marked by the teacher, so there is no mechanistic relationship between test scores and teachers' perceptions.

Black, S. (2000), ‘Together again: The practice of looping keeps students with the same teachers’, *American School Board Journal* **187**(6), 40–43.

Bogart, V. S. (2002), The Effects of Looping on the Academic Achievement of Elementary School Students, PhD thesis, East Tennessee State University.

Bryk, A. S., Sebring, P. B., Allensworth, E., Easton, J. Q. & Luppescu, S. (2010), *Organizing schools for improvement: Lessons from Chicago*, University of Chicago Press.

Burke, D. L. (1996), ‘Multi-year teacher/student relationships are a long-overdue arrangement’, *Phi Delta Kappan* **77**(5), 360.

Calonico, S., Cattaneo, M. D. & Titiunik, R. (2014), ‘Robust nonparametric confidence intervals for regression-discontinuity designs’, *Econometrica* **82**(6), 2295–2326.

Cattaneo, M. D., Jansson, M. & Ma, X. (2020), ‘Simple local polynomial density estimators’, *Journal of the American Statistical Association* **115**(531), 1449–1455.

Chetty, R., Friedman, J. N. & Rockoff, J. E. (2014), ‘Measuring the impacts of teachers I: Evaluating bias in teacher value-added estimates’, *American Economic Review* **104**(9), 2593–2632.

Cistone, P. & Shneyderman, A. (2004), ‘Looping: An empirical evaluation.’, *International Journal of Educational Policy, Research, and Practice: Reconceptualizing Childhood Studies* **5**(1), 47–61.

Clotfelter, C. T., Ladd, H. F. & Vigdor, J. (2005), ‘Who teaches whom? Race and the distribution of novice teachers’, *Economics of Education review* **24**(4), 377–392.

Clotfelter, C. T., Ladd, H. F. & Vigdor, J. L. (2006), ‘Teacher-student matching and the assessment of teacher effectiveness’, *Journal of Human Resources* **41**(4), 778–820.

Clotfelter, C. T., Ladd, H. F. & Vigdor, J. L. (2010), 'Teacher credentials and student achievement in high school a cross-subject analysis with student fixed effects', *Journal of Human Resources* **45**(3), 655–681.

Cohen-Vogel, L. (2011), '"Staffing to the test": are today's school personnel practices evidence based?', *Educational Evaluation and Policy Analysis* **33**(4), 483–505.

Comi, S. L., Argentin, G., Gui, M., Origo, F. & Pagani, L. (2017), 'Is it the way they use it? Teachers, ICT and student achievement', *Economics of Education Review* **56**, 24–39.

Contreras, D. (2019), Gender differences in grading: teacher bias or student behaviour? Mimeo, University of Nottingham.

Correia, S. (2016), Linear models with high-dimensional fixed effects: An efficient and feasible estimator, Technical report. Working Paper.

Cuesta, J. I., González, F. & Philippi, C. L. (2020), 'Distorted quality signals in school markets', *Journal of Development Economics* **147**, 102532.

Dee, T. S. (2007), 'Teachers and the gender gaps in student achievement', *Journal of Human Resources* **42**(3), 528–554.

Feng, L. (2010), 'Hire today, gone tomorrow: new teacher classroom assignments and teacher mobility', *Education Finance and Policy* **5**(3), 278–316.

Fitzpatrick, M. D. & Lovenheim, M. F. (2014), 'Early retirement incentives and student achievement', *American Economic Journal: Economic Policy* **6**(3), 120–54.

Franz, D. P., Thompson, N. L., Fuller, B., Hare, R. D., Miller, N. C. & Walker, J. (2010), 'Evaluating mathematics achievement of middle school students in a looping environment', *School Science and Mathematics* **110**(6), 298–308.

Fryer, R. G. J. (2018), 'The "pupil" factory: specialization and the production of human capital in schools', *American Economic Review* **108**(3), 616–56.

Graham, B. S., Ridder, G., Thiemann, P. M. & Zamarro, G. (2020), Teacher-to-classroom assignment and student achievement. NBER Working Paper No. 27543.

Guimaraes, P. & Portugal, P. (2010), 'A simple feasible procedure to fit models with high-dimensional fixed effects', *The Stata Journal* **10**(4), 628–649.

Hanushek, E. A., Kain, J. F. & Rivkin, S. G. (2004), 'Why public schools lose teachers', *Journal of Human Resources* **39**(2), 326–354.

Harris, D. N. & Sass, T. R. (2011), 'Teacher training, teacher quality and student achievement', *Journal of Public Economics* **95**(7), 798 – 812.

Hill, A. J. & Jones, D. B. (2018), 'A teacher who knows me: The academic benefits of repeat student-teacher matches', *Economics of Education Review* **64**, 1–12.

Imbens, G. W. & Lemieux, T. (2008), 'Regression discontinuity designs: A guide to practice', *Journal of Econometrics* **142**(2), 615–635.

Kalogrides, D., Loeb, S. & Béteille, T. (2013), 'Systematic sorting: Teacher characteristics and class assignments', *Sociology of Education* **86**(2), 103–123.

Kerr, D. L. (2002), "In the loop": responses about looping at the middle school level as seen through different lenses, PhD thesis, National-Louis University.

Klugman, J. (2017), 'Essential or expendable supports? assessing the relationship between school climate and student outcomes', *Sociological Science* **4**, 31–53.

Kraft, M. A., Marinell, W. H. & Shen-Wei Yee, D. (2016), 'School organizational contexts, teacher turnover, and student achievement: evidence from panel data', *American Educational Research Journal* **53**(5), 1411–1449.

Lavy, V. (2015), 'Do differences in schools' instruction time explain international achievement gaps? Evidence from developed and developing countries', *The Economic Journal* **125**(588), F397–F424.

Little, T. S. & Dacus, N. B. (1999), 'Looping: Moving up with the class.', *Educational Leadership* **57**(1), 42–45.

Liu, J.-Q. (1997), 'The emotional bond between teachers and students: Multi-year relationships', *Phi Delta Kappan* **79**(2), 156.

Nichols, J. D. & Nichols, G. W. (2002), 'The impact of looping and non-looping classroom environments on parental attitudes', *Educational Research Quarterly* **26**(1), 23.

Osborne-Lampkin, L. & Cohen-Vogel, L. (2014), '"Spreading the wealth": how principals use performance data to populate classrooms', *Leadership and Policy in Schools* **13**(2), 188–208.

Ost, B. (2014), 'How do teachers improve? the relative importance of specific and general human capital', *American Economic Journal: Applied Economics* **6**(2), 127–51.

Paredes, V. (2014), 'A teacher like me or a student like me? Role model versus teacher bias effect', *Economics of Education Review* **39**, 38–49.

Rivkin, S. G., Hanushek, E. A. & Kain, J. F. (2005), 'Teachers, schools, and academic achievement', *Econometrica* **73**(2), 417–458.

Rockoff, J. E. (2004), 'The impact of individual teachers on student achievement: evidence from panel data', *American Economic Review* **94**(2), 247–252.

Santiago, P., Fiszbein, A., Jaramillo, S. G. & Radinger, T. (2017), *OECD Reviews of School Resources: Chile 2017*, OECD publishing.

Thapa, A., Cohen, J., Guffey, S. & Higgins-D'Alessandro, A. (2013), 'A review of school climate research', *Review of Educational Research* **83**(3), 357–385.

Tourigny, R., Plante, I. & Raby, C. (2019), 'Do students in a looping classroom get higher grades and report a better teacher-student relationship than those in a traditional setting?', *Educational Studies* pp. 1–16.

Tucker, S. C. (2006), The impact of looping on student achievement on the Colorado student assessment program, Master's thesis, Regis University.

Zahorik, J. A. & Dichanz, H. (1994), 'Teaching for understanding in german schools.', *Educational Leadership* **51**(5), 75–77.

Appendix A School panel

In this Appendix we provide evidence on the extent to which looping can be considered a school-level policy. We use data on students between grade 5 and grade 8, during the years 2002-2018. Then we link students with their classroom teacher in four different subjects (Spanish, maths, social sciences and natural sciences). Using this information, we identify repeat matches in grade 6, grade 7 and grade 8. As a result, we obtain a sample of 12,102,819 students, 8,379 schools and 444,859 classes.

We aggregate this data to the school-year level and calculate the proportion of repeat matches across all subjects and grades, $\bar{R}_{kt}$. At this level, the sample contains 116,812 observations on 8,379 schools. A variance decomposition exercise reveals that the variation in looping within schools (overtime) is almost exactly equal to the variation in average looping behaviour between schools. Figure A1 shows the distribution of $\bar{R}_k$, which indicates that very few schools always or never use repeat matches.

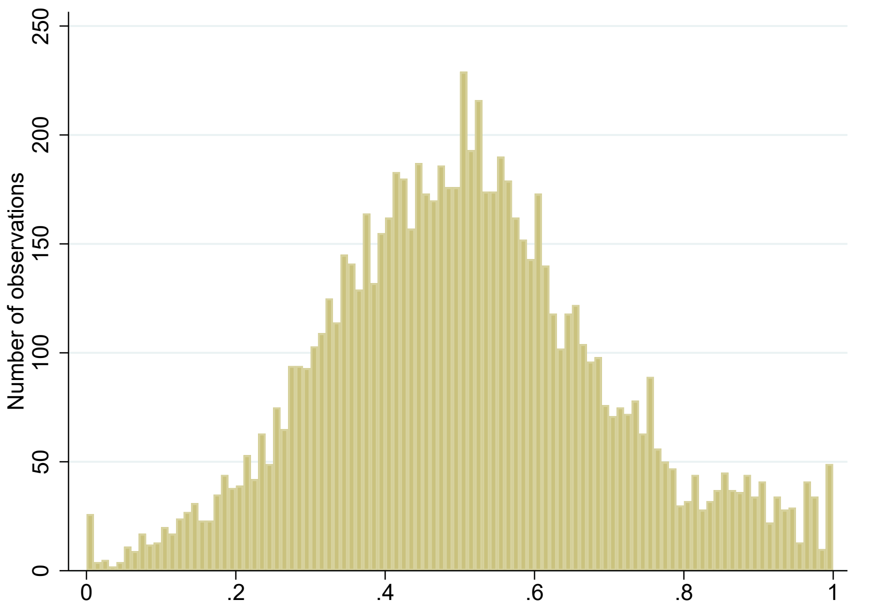


Figure A1. Distribution of the proportion of repeat matches at school level

We then aggregate the data to the school-subject-grade-year level, and again compute the proportion of repeat matches, $\bar{R}_{ksgt}$. To quantify how much of the variation in looping can be attributed to schools we estimate the following specification:

$$\bar{R}_{ksgt} = \mu_k + \mu_s + \mu_g + \mu_t \tag{7}$$

where μ_k is a school fixed effect, μ_s is a subject fixed effect, μ_g is a grade fixed effect and μ_t is year fixed effect. Column 1 and Column 2 in Table A1 show the benchmark model without and with school fixed effects, respectively. The results show only small variation in repeat matches across subjects and rather larger effects across grades. The inclusion of school fixed effects increases the adjusted R^2 from 2% to only 15%, from which we conclude that the prevalence of looping is only weakly associated with school-level decisions.

Table A1. Contribution of school fixed effects to the proportion of repeat matches

	(1)	(2)
1=Math	0.023^{***}	0.023^{***}
1=Natural	0.020^{***}	0.020^{***}
1=Social	0.014^{***}	0.014^{***}
1=Grade 7	-0.120^{***}	-0.071^{***}
1=Grade 8	0.018^{***}	0.068^{***}
School FE		Yes
Year FE	Yes	Yes
Adjusted R-squared	0.021	0.147
Observations	1,204,200	1,204,200

Notes: Dependent variable is the proportion of repeat matches at school-subject-grade-year level. $^{*}p < 0.10$, $^{**}p < 0.05$, $^{***}p < 0.01$.

Appendix B The exogeneity of the LRA discontinuity

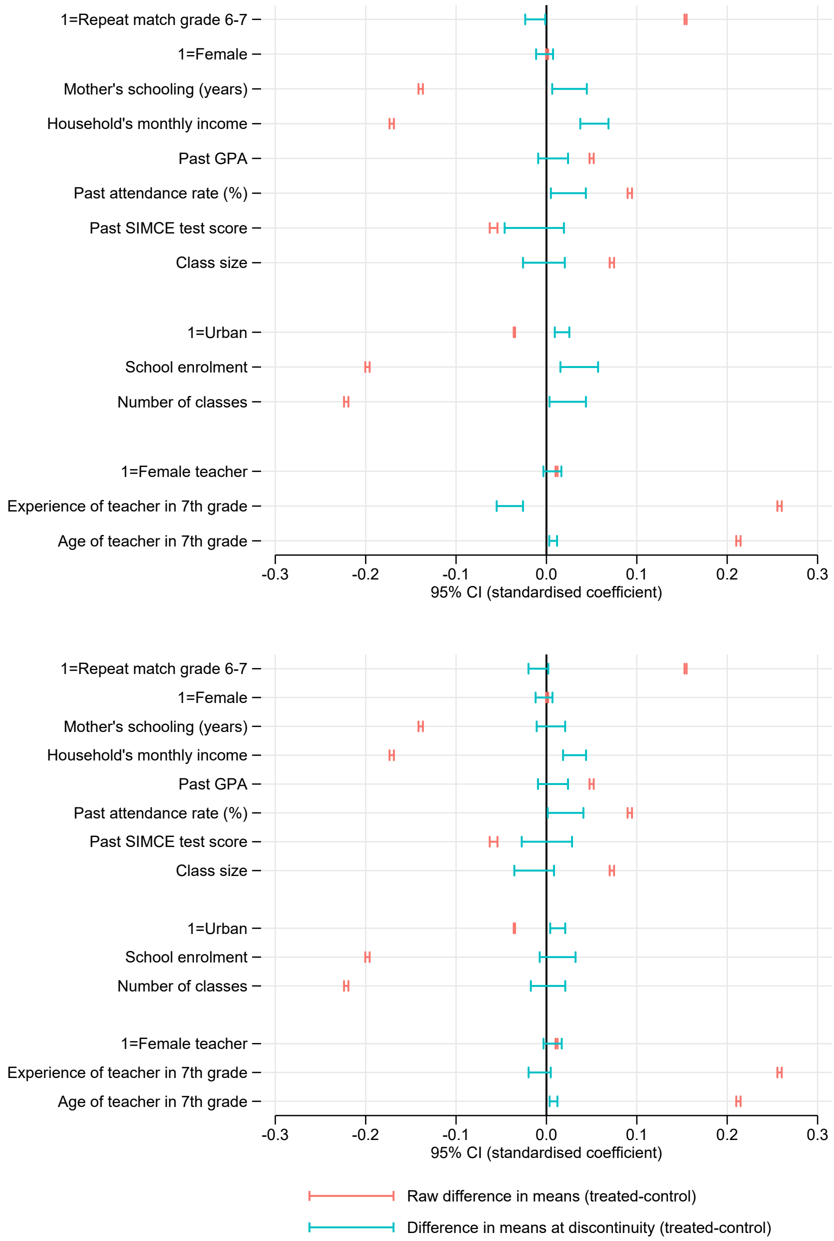


Figure B1. Balancing tests at the discontinuity

Notes: Figures show 95% confidence intervals on the difference in means between the treated and controls in the overall sample and at the discontinuity in the LRA. All variables are standardised to have zero mean and unit standard deviation to enable comparison. The bottom panel includes as covariates dummies for school type (Public, Private, Voucher). The difference at the discontinuity is estimated using methodology proposed by Calonico et al. (2014), with a polynomial of order one and weighted by triangular kernel. The number of observations for all the regressions is 2,785,928. Standard errors calculated using Calonico et al. (2014) and clustered at the student level.

Table B1. Tests of balance of income bands at the discontinuity, controlling by school type

Covariate	*RD estimator* (1)	*Std. err.* (2)	*Obs. left* (3)	*Obs. right* (4)	*Bandwidth* (5)
Income level: 1	0.001	0.004	232,241	116,886	1098.770
Income level: 2	0.004	0.004	234,639	117,513	1110.402
Income level: 3	0.002	0.004	243,495	119,672	1152.777
Income level: 4	−0.004	0.003	205,880	110,872	987.444
Income level: 5	−0.000	0.002	269,131	124,478	1267.990
Income level: 6	−0.002	0.002	241,984	119,275	1143.496
Income level: 7	−0.000	0.001	234,098	117,224	1107.541
Income level: 8	−0.001	0.001	206,427	110,942	990.439
Income level: 9	−0.000	0.001	257,206	122,031	1217.552
Income level: 10	0.000	0.001	300,977	128,863	1387.952
Income level: 11	0.000	0.001	293,698	128,187	1357.590
Income level: 12	−0.000	0.000	237,756	118,151	1127.060
Income level: 13	−0.000	0.000	258,257	122,226	1222.284
Income level: 14	−0.000	0.000	313,007	130,989	1436.014
Income level: 15	−0.001	0.001	123,887	84,820	619.084

Notes: Table shows the estimated discontinuity in each income band at the LRA controlling for school type. Results based on the empirical strategy that implements a RD following the methodology proposed by Calonico et al. (2014), with a polynomial of order one and weighted by triangular kernel. The number of observations for all the regressions is 2,785,928. Standard errors calculated using Calonico et al. (2014) and clustered at the student level.

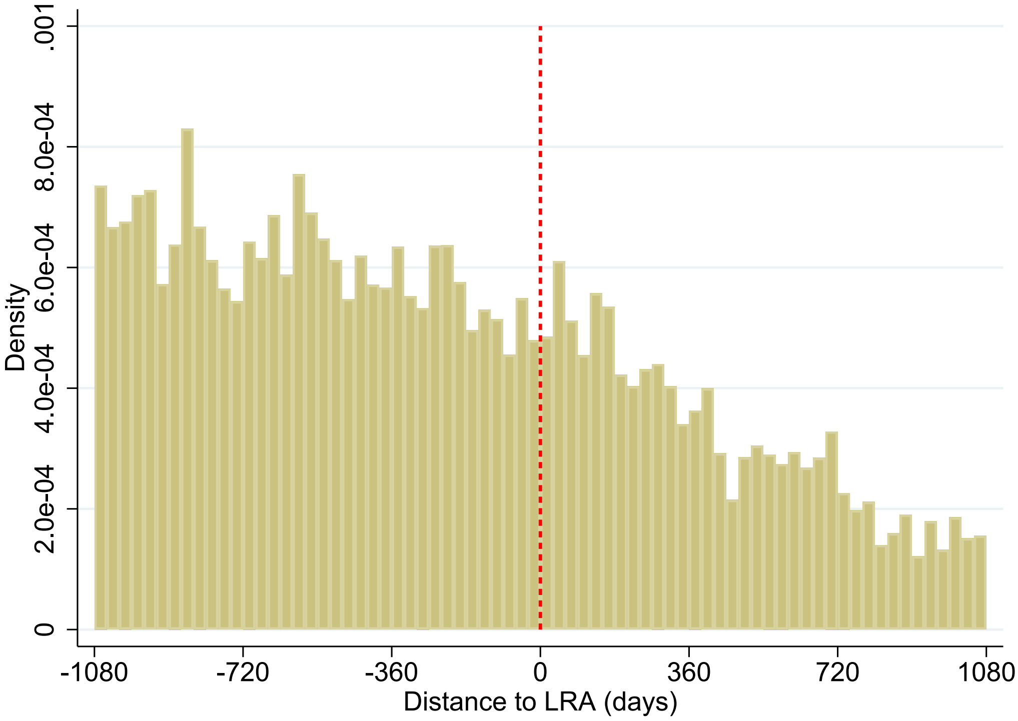


Figure B2. Density of the running variable

Notes: Running variable is distance, in days, from age on the final day of the grade 7 school year to the day on which the teacher reaches the legal retirement age. Bins have width of 30 days.

Appendix C Experience Model

Table C1. Returns to experience across different experience ranges

	(1)
1–2 years of experience	0.014***
	(0.003)
3–4 years of experience	0.032***
	(0.003)
5–9 years of experience	0.037***
	(0.003)
10–14 years of experience	0.040***
	(0.003)
15–24 years of experience	0.036***
	(0.003)
>25 years of experience	0.024***
	(0.003)
Student FE	Yes
Subject FE	Yes
R-squared	0.793
Observations	2,785,928

Notes: Dependent variable is the student's SIMCE test score in grade 8. The model shows the estimated returns to experience across different experience ranges. The omitted category is teachers with zero experience. The model includes a female teacher dummy. Standard errors are clustered at the student level. $^{*}p < 0.10$, $^{**}p < 0.05$, $^{***}p < 0.01$.

Appendix D Regression discontinuity results by subject

Table D1. Effect of repeated matches on test scores by subject: linear regression discontinuity results

	Spanish *(1)*	*Maths* *(2)*	*Natural Sciences* *(3)*	*Social Sciences* *(4)*
τ_R (First stage)	-0.088^{***}	-0.164^{***}	-0.111^{***}	-0.161^{***}
	(0.003)	(0.004)	(0.004)	(0.004)
τ_y (Reduced form)	-0.009	-0.013	-0.005	-0.014
	(0.010)	(0.012)	(0.012)	(0.011)
τ_{RD}	0.102	0.080	0.049	0.085
	(0.118)	(0.073)	(0.105)	(0.071)
Teacher FE $\times$ Year FE	Yes	Yes	Yes	Yes
First-stage R-squared	0.762	0.753	0.760	0.765
First-stage F statistic	654	1,498	780	1,826
Observations	696,482	696,482	696,482	696,482

Notes: Dependent variable is the student's SIMCE test score in grade 8 in each subject. Treatment is the student-subject measure of repeated match R_{is}. All the models therefore control for student characteristics (gender, household income, mother's education, final GPA and lagged attendance rate), class size and school characteristics (public school dummy and rural school indicator). $^{*}p < 0.10$, $^{**}p < 0.05$, $^{***}p < 0.01$.

Appendix E Information from teachers' survey

Table E1. Mean comparison test of classroom characteristics, full sample versus estimation sample

	Estimation sample	Sample with teachers' perception	Difference	Std. err.
1=Female	0.50	0.50	-0.005^{**}	(0.002)
Mother's schooling (years)	10.67	11.12	-0.443^{***}	(0.032)
Household's monthly income	378.51	437.00	-58.486^{***}	(5.033)
Average SIMCE test score	-0.06	-0.03	-0.028^{***}	(0.007)
Past GPA	0.08	0.09	-0.017^{***}	(0.004)
Past attendance rate	94.34	94.08	0.257^{***}	(0.041)
Class size	21.88	21.02	0.855^{***}	(0.120)
1=Public	0.53	0.44	0.084^{***}	(0.006)
1=Urban	0.82	0.82	-0.008^{*}	(0.005)
Observations	31,837	9,498		

Notes: $^{*}p < 0.10$, $^{**}p < 0.05$, $^{***}p < 0.01$.