Do higher salaries yield better teachers and better student outcomes?

José María Cabrera* and Dinand Webbink**

Abstract

We study the effects of a policy aimed at attracting more experienced and better qualified teachers in primary schools in disadvantaged neighborhoods in Uruguay. Teachers in these schools could earn higher salaries, and more experienced teachers are given priority in choosing teaching positions. Eligibility for the program was based on a poverty index with a cutoff rule. Estimates from regression discontinuity models show that the policy successfully led to 'hiring experience from other schools', and also increased tenure. Overall, the effect on student outcomes was small. We rationalize this result by showing that the program may have increased experience in ways that are not strongly associated with improved student outcomes. Consistent with this, we do find achievement gains for students in schools that saw a reduction in the share of very inexperienced teachers. The results underscore that increases in teacher pay may only improve student outcomes if it increases those teacher characteristics that actually improve student outcomes.

JEL Codes: I2, J24

Keywords: teacher salaries, teacher experience, student performance, disadvantaged students.

* Universidad de Montevideo, Uruguay

** Erasmus School of Economics, Rotterdam, Tinbergen Institute, IZA,
Corresponding author: webbink@ese.eur.nl

We are grateful to both referees and the editor Kirabo Jackson for all their comments and suggestions. We also thank Juan Dubra, Anne Gielen, Sacha Kapoor, Olivier Marie, Matthijs Oosterveen, and seminar participants at the University of Rotterdam, Universidad de Montevideo, University of Tilburg, University of Padova, Bocconi University Milan, IECON, EALE 2015, ESPE 2015, XXIX Jornadas BCU, 2014 annual meeting of the Uruguayan Economic Association, workshop The Hague, Maastricht workshop 2016 for helpful comments. Financial support from ANII (POS_EXT_2012_1_9992) to José María Cabrera is appreciated.

1. Introduction

In general, more experienced and better qualified teachers are less likely to teach in schools that serve children from relatively poor families (e.g. Hanushek et al. 2004; Clotfelter et al. 2007, 2008; Jackson 2009). This allocation of teachers across schools might have important short term and long term consequences as there are large differences between teachers in their impact on student outcomes (Rockoff, 2004; Rivkin et al., 2005; Aaronson et al., 2007; Kane & Staiger, 2008; Hanushek & Rivkin, 2010; Chetty et al., 2014; Hanushek, 2011). In particular, differences in experience among teachers, especially in the initial years of the teaching career, are related to differences in student outcomes (Rivkin et al. 2005; Rockoff 2004; Harris & Sass 2011; Papay & Kraft 2015; Wiswall 2013). As such high, concentrations of rookie teachers in certain schools can be particularly deleterious to students at those schools. To address these issues one might consider offering higher salaries with the aim of attracting more experienced and better qualified teachers for working in schools in disadvantaged neighborhoods. This basic policy has been proposed in many settings around the world.

This study investigates the impact of a policy program, the so-called CSCC-program, which explicitly aimed to reallocate more experienced and better qualified teachers to schools in poor neighborhoods. Teachers in primary education in Uruguay could obtain increases in their base salaries up to 26 % by working in schools in poor neighborhoods[1]. This policy was expected to have an impact especially on the decisions of experienced teachers due to specific characteristics of the teacher labor market. Teachers in primary education in Uruguay have to apply for a job each year. The application process is administered by a central authority which sets the criteria for the assignment of teachers to jobs in schools. The key feature of the system is that experienced teachers may choose first from the available teaching slots among schools. This implies that more experienced teachers are given priority in the decision at which school to work and in the decision about obtaining a higher salary through working in the poor neighborhoods. In this study we investigate two potential impacts of this funding policy. First, we investigate the impact of the policy on the experience and tenure of the teaching staff in schools in the targeted

[1] The program also includes several non-salary components, like equipment and learning materials, but these components appear to be less important, see Section 2.

neighborhoods. Second, we study the impact of the policy on test scores and non-test-score outcomes of students in poor neighborhoods.

For identifying the impact of the policy we exploit variation between schools induced by the eligibility rule for participation of schools in the program. Since 2005 eligibility for the program is determined by a poverty index that aims to capture the living conditions in neighborhoods. Schools with a score on the poverty index above a certain threshold are eligible for the program. This assignment scheme allows the use of regression discontinuity models for estimating intention to treat and treatment on the treated effects of the program. For the estimation we use administrative panel data of all public primary schools in Uruguay since 1992. This database contains information about the experience and tenure of teachers at the grade level of all public schools[2]. Moreover, we use data about test scores and non-test score outcomes (grade retention, dropout and insufficient attendance). In general, test scores have been the focus of economic research about educational outcomes but they might miss changes in 'non-cognitive skills' that have been shown to be important for adult outcomes[3]. A recent literature relates non-test-score outcomes to character skills, and shows that teachers can have meaningful effects on non-test-score outcomes (Jackson, 2018; Gershenson, 2016; Ladd & Sorenson, 2017). Our data allow us to study the effects of the program on these 'cognitive and non-cognitive' outcomes up to eight years after 2005. This implies that we can yield estimates that are robust to implementation issues from the initial years of the redesigning of the program.

This paper investigates the effects of an unconditional pay increase within a teacher labor market setting that favors more experienced teachers. In general, it is not clear whether an unconditional pay increase may have an impact on student performance. A pay raise will only improve student outcomes if it increases those teacher attributes that lead to better student outcomes (e.g. Jackson, Johnson and Persico, 2016). For example, a pay raise that leads to more teachers with Master's degree (which has been found to have no effect on student outcomes) may

[2] On average schools have two teachers per grade. The results are very similar when we use data of individual teachers from our database.

[3] See e.g. Borghans, Weel and Weinberg (2008), Heckman & Rubinstein (2001), Heckman et al. (2006, 2013) and Lindqvist and Vestman (2011). Several studies also find that long term outcomes of educational interventions cannot be fully explained by changes in test scores (Chetty et al. 2011; Fredriksson et al. 2013; Heckman et al. 2013).

have no impact. However, a pay raise that replaces the highly inexperienced teachers with experienced teachers may improve student outcomes. If schools in targeted neighborhoods will be able to obtain a more experienced teaching staff, then the impact of the policy on student outcomes will depend on whether these teachers are more productive. The empirical literature points to two aspects of experience that might be important for teacher productivity.

First, it is well established that the gains to teacher experience are most pronounced in the first few years (e.g. Hendricks 2014). As such, replacing mid-career teachers (5 -10 years) with veteran teachers (10 or more) may have little impact since they are similarly effective. However, replacing early career teachers (0 to 5) years with mid-career teachers (5 -10 years) could yield considerable gains.

Second, experience can be obtained within the current school but also within other schools, and hence will contain general and specific components of human capital. The effects of these components on student outcomes might differ (Jackson 2013; Ost 2014). Moreover, hiring experience from other schools increases turnover which will impact student outcomes. As such, the impact of the policy might depend on whether it especially affects 'hiring of new teachers' or 'keeping of current teachers'.

We find that the program is successful in increasing experience on average, especially by reallocating experienced teachers to schools in the targeted neighborhoods. Estimates from regression discontinuity models show large effects on the teaching staff in the targeted schools. Since 2009 the policy on average increased the experience of the teaching staff by two to three years. This means that in each class within each grade the experience of the teacher increased on average with two to three years. For schools that actually participated in the program (the treatment on the treated effect) we find an increase of experience with three to seven years. Moreover, we also find that tenure (years of experience at the current school) increased by approximately one year in the targeted schools, and by one to two years in the schools that participated in the program. This difference in the experience of the teaching staff between schools across the cut-off for program eligibility stayed on for several years.

Despite these large changes in the composition of the teaching staff the impact on student performance appears to be small. Across a range of non-test-score measures and test-score outcomes we fail to detect a consistent positive impact of the CSCC-program. We find some evidence that the program increases school attendance of students, especially the attendance of students in grades 1 and 2. However, these results are not robust to the specification of the forcing variable or to the discontinuity sample that is used for the estimation. We don't find an effect of the program on grade retention or drop out, nor on test scores in math and language. Student mobility is unlikely to affect the estimated effects due to the application of fixed catchment areas in primary education in Uruguay. Moreover, we don't find an increase of the proportion of disadvantaged students in the targeted schools. This leads us to conclude that the effects of the CSCC-program on student outcomes are likely to be quite small.

The non-linearity of the teacher productivity-experience profiles (Hendricks, 2014; Papay & Kraft, 2015) appears to be important for explaining the modest results. The program induces replacement of teachers who are on the steepest part of the productivity-experience profile, but also replaces teachers who are on parts of the productivity-profile that are less steep. The latter replacements are less efficient, and thereby might reduce the impact of the program. Consistent with this, we find a clear difference between schools that had many rookie teachers and schools that had little rookie teachers at the start of the reform. The program improved student outcomes in schools with many rookie teachers but didn't improve outcomes in schools with little rookie teachers.

We also find that the distinction between tenure and experience ('hiring' and 'keeping' teachers) might be important for explaining the modest results of the program. We find that the program especially affects 'hired experience'. Moreover, 'hired experience' is not associated with student outcomes, whereas tenure (years of experience in the current school) is found to be important for student outcomes. This implies that the CSCC-program especially affects the component of experience that is less relevant for student performance.

Our study contributes to several branches of the literature that studies the relationship between the teacher labor market and school performance[4]. The first part of our paper is related to the literature that investigates the effect of teacher pay on turnover rates (see for instance Dolton and Van der Klaauw 1995, 1999; Murnane et al. 1989; Hanushek et al. 2004; Imazeki 2005; Clotfelter et al. 2011; Gilpin 2011). These studies typically find that increases in teacher pay reduce turnover of teachers. However, these findings might be biased as changes in teacher

[4] This paper focuses on the teacher labor market. For a comprehensive review of the literature on selection and incentive structures of public sector workers, see Finan et al. (2015).

pay might be correlated with unobserved time-varying characteristics. The estimated effects tend to be smaller in studies that use arguably exogenous variation (Clotfelter et al. 2008; Hendricks 2014). Our study contributes to this literature by using exogenous variation in teacher pay and focusing on the changes in the teaching staff that result from changes in tenure and from the entry of teachers that obtained experience at other schools.

The second part of our paper is related to the literature that studies the relationship between teacher pay and student outcomes. Numerous studies have investigated this relationship but the evidence appears to be mixed. For instance, a review study about the effects of school inputs on student performance based on 90 studies (Hanushek, 1997, 2003) reports that only 20 % of 119 estimates finds a positive effect of teacher pay on school performance. Several recent studies suggest that teacher pay has a positive effect on student performance (Loeb & Page 2000; Britton & Propper 2016; Dolton & Marcenaro-Gutierrez 2011). However, a recent study based on a randomized experiment finds no effect of the unconditional doubling of teacher pay on student performance after two and three years (De Ree et al. 2018). Our study contributes to this literature by looking at effects of higher teacher pay over a period of six years and using a transparent identification approach. A longer period for investigating the program effects is important as it is not clear how and when teacher pay will affect student performance (Hendricks, 2014). It should be noted that the context of our study differs from previous studies. We investigate the impact of teacher pay within an education system that gives priority to more experienced teachers.

This study also contributes to the branch of the literature that looks at the importance of teacher characteristics, in particular teacher experience, for student performance (e.g. Rivkin et al. 2005; Rockoff 2004; Harris & Sass 2011; Papay & Kraft 2015; Wiswall 2013; Gerritsen et al. 2017). These studies show that teacher experience matters, especially in the first four or five years of the teaching career. Hendricks (2014) exploits these typical productivity-experience profiles for linking retention rates of teachers induced by changes in teacher pay to school performance. He finds that teacher pay reduces turnover rates but the effects on school performance are small. Our study contributes to this literature by investigating the impact of higher teacher salaries on changes in the teaching staff and on student outcomes for the same schools. Moreover, in our study we can distinguish between changes in the teaching staff due to changes in tenure and due to newly hired teachers.

The remainder of this paper is organized as follows. Section 2 describes the institutional background and the specific program that we study in this paper. Section 3 discusses the data used in the estimation. In Section 4 we explain the empirical strategy. Sections 5 and 6 show the main estimation results about the impact of the program on teachers and on students. In Section 7 we focus on the mechanisms that might explain our main finding; the modest effects on student outcomes despite the large increases in teacher experience. Finally, Section 8 concludes.

2. Institutional Background and the CSCC-Program

Uruguay is a country of 3.4 million inhabitants located in the south-eastern region of South America, sharing borders with Argentina and Brazil. It has a per capita income of USD 21,944 which places Uruguay in the World Bank list of high-income economies. In the PISA 2012 test Uruguayan students scored 409 points, which places them high at the regional level but low compared to the OECD average (494 points). The educational system delivers quite unequal results. For instance, it has one of the largest standard deviations in the performance of schools and between 2003 and 2012 inequality has further increased.

Primary education in Uruguay has universal coverage and the majority of children attend schools from the public system (85% in 2010). Students attend two years of preschool education *(Educación Inicial)* and six years of primary school (*Escuela Primaria).* Most students enrol in so-called Standard Urban schools. These schools have approximately 350 students (50 in preschool and 300 in primary school), 11 teachers and a class size of 25 students. The program that we will study in this paper provides extra resources to *Standard Urban* schools in poor neighbourhoods. The central administration of the education system (central authority) plays a key role in the system. This agency is responsible for the hiring of teachers and the payment of their salaries. Schools don’t hire teachers themselves but this is done by the central authority. Only certified teachers are allowed to work in primary education.

The CSCC- program

The CSCC- program (*Contexto Socio Cultural Crítico*) provides extra resources to schools that are located in disadvantaged areas of the country. The aim of the program is to improve school performance of children from families living in poor neighbourhoods. In Uruguay students are obliged to attend the school in their neighbourhood. The program has been implemented since 1995 at the national level, and there are *CSCC* schools in every state of the country. In 1995 the first 155 schools were assigned to the program based on indicators of school performance and characteristics of the neighbourhood. Since 1995 the number of participating schools and the eligibility rules of the program have changed several times mostly due to changes in the available governmental resources (see Appendix A).

The CSCC program consists of several components. The main, and the only component of the program that is strictly enforced, is an increase in the salary of teachers. The higher salaries for teachers that decide to work in the CSCC schools have the explicit aim of getting a more experienced and better qualified teaching staff in the poor schools. The following quote taken from the Parliamentary Budget Law (2005) illustrates this: “The higher salary will be offered to teachers working at CSCC schools, with the aim of encouraging to attract the most experienced and best qualified ones”.[5] The extra salary consists of a fixed amount for all teachers working in CSCC schools, and is not tied to performance or to a teacher’s position in the payment scale. The extra payment is determined as a 26% increase over the base salary in the first category of the payment scale. A teacher´s full salary consists of the base salary plus additional payments, which increase the base salary by approximately 70%.[6] The extra salary from the CSCC-program increases the full salary for a teacher in the first level of the payment scale by approximately 15%. Table A.1 in the appendix shows teachers’ salaries from 1997 to 2010 by payment scale.

[5]The original text in Spanish: “El mayor salario también atenderá a los docentes que cumplan sus funciones en los centros educativos de contexto socio cultural crítico, <u>con el objetivo de propiciar la radicación de los más experimentados y mejor calificados</u>.”

[6] The calculation of teacher salaries is quite complicated. For instance, the additional payments for teachers consist of 20 items.

In this study, we use the salary variation caused by CSCC to better understand the relationship between teacher pay, teacher quality and student outcomes. However, the CSCC, in principal has a few additional complements that we agree are not at play. The CSCC-program also provides schools with equipment, learning materials and improved lunchrooms. Moreover, the program includes components like ‘additional time for coordination between teachers’ (aimed at institutional development activities, curriculum planning, coordination on program content and evaluation criteria, etc.) and (voluntary) ‘training sessions for teachers’. These components of the program are similar to the ones implemented in other countries, for instance in Chile (Chay, McEwan and Urquiola, 2005). The importance of these non-salary components is not clear as they are not strictly enforced, and there are no statistics available about the allocation of the budget of the CSCC program.[7] Statistics are only available about the total budget of primary education; they show that more than 80 % is spent on teacher salaries.[8] Nevertheless, our database might shed some light on the importance of the non-salary components as it contains information about several school resources that are quite similar to the non-salary components of the CSCC program (% children having lunch at school, school has library room, computers for educational use, number of computers, number of study books per pupil in 1st grade, school has community teacher). We have investigated whether the schools that we are comparing in our evaluation of the effects of the CSCC-program differ with respect to the use of these resources. For this analysis we have used the main regression discontinuity models that will be explained in Section 4. The estimates from these models, which are shown in Table A.2 in the appendix, reveal no difference in the use of these resources. It should also be noted that the information on these resources is available for the years in which there was a high compliance with the eligibility rules of the program (see below). This implies that the treatment effects of the program that we are estimating in the next sections are mainly driven by the salary component.

Eligibility for the program is completely determined by the central administration of the education system. Schools don’t apply for the program, nor can they opt in or out of the program.

[7] From the accountants of the central administration we learned that CSCC schools were given priority with respect to the allocation of the other components. In case that there were not enough resources for all schools then the resources were first delivered at the CSCC schools.

[8] See: http://www.anep.edu.uy/anep/phocadownload/Publicaciones_DSPP/gastos%20y%20salarios%201985-2011.pdf

This bureaucratic feature of the program reduces concerns about non-random selection of schools into the program. Key to our analysis is the redesigning of the program in 2005. Since 2005 eligibility for the program is determined by a poverty index. This index is based on a set of indicators about parental education, poverty and social integration and created by using factor analysis[9]. Schools were ranked according to their score on the index, and eligibility for the program was determined by a threshold value. The redesigning of the program in 2005 provides a transparent rule for the assignment of schools to the program which is important for the identification of the effects. The poverty score of 2005 also determined eligibility for the program for the following years until 2010. Hence, the poverty score doesn't change between years. In our empirical approach we exploit the poverty index and the threshold value within a regression discontinuity framework (see Section 4). The CSCC-program already exists since 1995 but the eligibility rules for the period before 2005 are not clear. We, therefore, focus the analysis on the period since 2005. The eligibility rules changed again in 2011 and induced non-compliance with the eligibility rules of 2005.

Figure A.1 in the appendix provides information about the implementation of the redesigning of the program for the period 2005-2010. The figure shows participation in the program (Y-axis) by score on the poverty index. Schools with poverty scores above zero are eligible for the program. In 2009 and 2010 all schools that were eligible for the program actually participated in the program. Moreover, there were no schools participating that were not eligible. This means that there is full-compliance with the assignment rules in 2009 and 2010. In the years before we observe a transition towards full compliance. Figure A.1 also shows that this transition started with the inclusion of schools in the poorest areas into the program, hence, schools further away from the cut-off were the first to be included in the program.

[9] Poverty was measured with an unmet basic needs index based on information about overcrowded homes, the materials used to construct the house, where families obtain water to drink, and the sanitary services of the house. The social integration index was constructed from information on integration in the territory (percentage of students living in illegal land), integration in the education system of brothers and sisters of the students, and integration of the household head into the labor market.

The CSCC-program and teacher experience

A special feature of the education system in Uruguay is that teachers have to apply each year for a school. The application process is run by the central administration of the education system. They provide a list of available teaching jobs for which teachers can apply. Teachers can choose from this list the school at which they would like to work. Hence, schools don't hire their teachers themselves, but teachers make the decision about the school. A key element in the application procedure is that teachers are ranked based on specific criteria set by the central administration. The experience of teachers is an important criterion for this ranking next to, for instance, pedagogical criteria[10]. The rank of the teacher determines the allocation of teachers to jobs. The teacher with the first rank is also the first who can choose from the available teaching slots[11]. Through this system more experienced teachers may choose first from the available teaching slots at schools. They will consider various factors such as commuting distance, school amenities, student population and school performance. This application procedure results in a negative correlation between the experience of the teaching staff and the poverty index of the school. For instance, in the year of the redesign of the policy (2005) this correlation is -0.25, which is statistically significant at the 1 %-level. This implies that students in poor neighbourhoods are taught by less experienced teachers.

The *CSCC Program* tries to mitigate this type of sorting by providing higher salaries to teachers that start working in the target schools. If the increased pay makes these schools more attractive, then they will be more likely to be chosen first by experienced teachers as they have first pick.

3. Data

The main data we use in this study come from an administrative registry the so-called Monitor Educativo de Enseñanza Primaria produced by the Department of Research and Statistics of the

[10] For more details see Article 13 of the following policy document: http://www.anep.edu.uy/anep/phocadownload/normativa/estatuto%20del%20funcionario%20docente_151130.pdf

[11] In cases of ties there are rules about other factors that should be taken into consideration such as seniority in a teaching category, time of entering the public system first or year of graduation.

National Administration of Public Education (ANEP). This is the official source of information on the public education system of the country. The database has been compiled since 1992 and has been produced in a standardized way over the years. It is based on regular administrative registries, annual questionnaires among school principals (since 2002), and surveys among parents. It includes information on the education process (enrolment, average group size, students per teacher), educational outcomes (insufficient attendance, repetition and dropouts), human resources at each school (number of teachers and other staff, teacher experience and tenure), material resources (library, lunchroom, other infrastructure, school equipment), and the social context of each school. This information is available for all public schools in the country. The database contains information at the school-level, at the grade-level and also information about individual teachers. The grade-level data are used in the main analysis. We use the individual teacher data in the analysis that looks at differences in the distribution of teachers within schools in Section 7.[12]

This study focuses on two types of outcomes. First, we look at the effect of the CSCC-program on the teaching staff of schools. For this analysis we use teacher experience (working years in education) and tenure (years of experience at the same school) as outcome variables. These measures are typically used in studies that investigate the importance of teachers for student performance (see Section 1). Both variables are measured in years and are available at the grade level. Hence, for each school we know the average experience and tenure at the grade level in each year. Information about experience and tenure is available since 2002.

The second type of outcomes is about student performance. The main database contains information about three non-test-score outcomes: insufficient attendance, grade retention and dropout. These non-test-score outcomes are very similar to the measures used in recent studies about the impact of teachers on ‘noncognitive skills’ and their longer-run outcomes (Jackson, forthcoming; Gershenson, 2016; Ladd & Sorenson, 2017). In our data insufficient attendance is defined as the percentage of students who attended school more than 70 days but less than 140 days in the academic year. Grade retention is the percentage of students that were retained in a

[12] Unfortunately, the data don’t contain teacher identifiers which could enable investigating individual movements of teachers caused by the program.

specific grade[13]. Decisions on grade retention are based on nationwide rules which, for instance, require a specific final score ('bueno') and 80 percent attendance during the school year (Circular 200/2008). These rules explicitly aim to standardize decisions regarding student progress in public primary education in the whole country[14]. Dropout is defined as the percentage of students that attended less than 70 days in the academic year. Information about these outcomes is available for each grade within each school since 1992. We have also constructed a fourth outcome variable 'attendance' based on more specific information about the number of days that students attended school. For each grade we have additional information about the number of students that attended school measured with the following categories: 1-70 days, 71-130 days, 131-140 days, 141-150 days, 151-160 days, 161-170 days, 171 days or more. This information allows us to construct a more detailed variable on school 'attendance'.

The main database doesn't contain test scores as there is no uniform test that is taken by all students in the country. However, for a sample of schools we were able to obtain achievement scores from standardized tests. The data come from the National Evaluation of Learning *(Evaluación Nacional de Aprendizajes),* which is the national assessment of learning, carried out by the central authority of the educational system. Math and language tests were taken by students in grade 6 in a representative sample of schools. We have data for the years 1999, 2002 and 2009, respectively from 190, 272 and 251 schools with approximately 7,500 students in each year.

The main database also contains a broad set of school and family characteristics. We use these variables for balancing tests around the cut-off. Table 1 shows summary statistics of the main variables in 2005 by eligibility for the program for all schools in the sample. In 2005 the database contains 374 schools at the left side of the cut-off and 284 schools at the right side of the cut-off. From these schools we have data about respectively 2,295 and 1,752 grades in 2005. We observe clear differences between eligible and non-eligible schools. Schools that are eligible for the program have a higher score on the poverty index and do worse on the family measures 'mother's education' and 'children with unmet basic needs'. Moreover, teachers working in

[13] Manacorda (2012) uses the same variable in his analysis of the effect of grade retention in grades 7 to 9.

[14] http://www.ceip.edu.uy/documentos/carpetaarchivos/normativa/circulares/Circular200_08.pdf

eligible schools on average have 12.2 years of experience, which is two years less than teachers at non-eligible schools. Teachers at eligible schools also have less tenure (the difference is more than one year). For the student outcomes we also observe clear differences. Eligible schools do worse on the three school performance outcomes and also experience more verbal and physical violence. Hence, schools that are eligible for the program are not a random draw from the population but score worse on teacher outcomes and student outcomes, and have less family resources. Changes in the outcome variables between 2005 and 2010 are shown in Table A.3 in the appendix.

4. Empirical strategy

In this paper we are interested in the effect of the CSCC-program on the school choice of teachers and on the performance of students. For identifying the effect of the program we exploit the redesigning of the program that was introduced in 2005. Since 2005 eligibility for the program is based on a poverty index with a clear cutoff rule. Schools with a poverty score above a certain threshold are eligible for treatment. Schools with poverty scores below this threshold are not eligible. This assignment scheme allows the use of a standard regression discontinuity model for estimating the causal effect of the program. In this approach we can use the poverty index as the forcing variable and estimate the following equation:

$$Y_{git} = \alpha_0 + \alpha_1 E_i + f(s_i) + \alpha_2 X_{git} + \varepsilon_{git} \qquad (1)$$

where Y_{git} is the outcome for grade g of school i at time t (e.g. experience of teaching staff at time t or student performance at time t), E_i is a dummy variable for eligibility for the CSCC program, s_i is the poverty score, X_{git} is a vector of control variables which also includes year and grade dummies, and ε_{git} are unobserved factors. In this specification $f(.)$ is a smooth function of the forcing variable which is allowed to be different at either side of the cutoff. Schools are eligible for treatment ($E_i = 1$) if their poverty score is equal to or above the threshold $s_i \geq s_o$. Estimation of the parameter α_1 will yield the causal effect of eligibility for the CSCC program since 2005 if there are no other discontinuities around the cutoff. This issue will be investigated below.

It is important to note that in our application program eligibility might affects schools on both sides of the cut-off. For instance, experienced teachers might move from non-eligible schools to schools that are eligible for the CSCC-program. Hence, due to the program some schools might gain experience and some schools might lose experience. The estimates of the parameter α_1 will therefore reflect these two potential effects of eligibility for the program.

In the first years after the redesigning of the program in 2005 there was non-compliance with the eligibility rule based on the poverty score. This non-compliance originates from two facts. First, the program already existed since 1995. Second, it took a few years to implement the new assignment rules of the program (see also Figure A.1). This explains that in the first years since 2005 some school that were not eligible still participated in the program and some school that were eligible didn't participate in the program. This non-compliance gradually reduced, and in 2009 and 2010 full compliance was achieved. In the years after 2010 the eligibility rules of the program changed again, which also induced non-compliance with the 2005-eligibility rules in 2011, 2012 and 2013. We estimate the treatment on the treated effect by using a standard IV-approach in which actual program participation (P_{it}) is instrumented with program eligibility (E_i):

$$P_{it} = \delta_0 + \delta_1 E_i + f(s_i) + \delta_2 X_{git} + v_{git} \quad (2)$$

$$Y_{git} = \beta_0 + \beta_1 P_{it} + f(s_i) + \beta_2 X_{git} + \eta_{git} \quad (3)$$

The estimate of the parameter β_1 can be interpreted as the causal effect of actual participation in the CSCC program on the outcomes in a specific year if the IV-assumptions with regard to the first stage, the independence assumption and the exclusion restriction hold. The instrument should be independent of the potential outcomes and the potential treatment statuses, and the instrument should only have an effect on the outcome through the endogenous variable (CSCC-program participation) (Imbens & Angrist, 1994).

Identification issues

The main assumption in the regression discontinuity model is that all observed and unobserved factors should behave smoothly around the cutoff. We have investigated this assumption by performing two tests. First, we looked at the density of the forcing variable at the cutoff to investigate whether schools might have manipulated their assignment to the program. It should be noted that this type of manipulation is not very likely as the assignment to the program is completely determined by the central authority (see Section 2). Second, we have performed balancing tests of covariates and outcomes in the baseline year (2005). The results of both tests don't raise concerns about manipulation of the assignment to program (see appendix B for a full description of the test and the results). In addition, we have checked for differences in program participation before 2005 across the cutoff as these differences might confound the estimates. We didn't find differences in program participation before 2005 (see appendix B). Another concern with our empirical approach is that there might have been changes in the composition of students in schools across the cutoff. We have tested whether the school size or the socioeconomic background of students has changed. These tests don't provide evidence that the main results might be affected by composition bias (see Appendix B).

Robustness analysis

For our main analyses we estimate models that include linear and quadratic specifications of the forcing variable. In addition, we perform a variety of robustness tests (see Appendix B). RD empirical results are often sensitive to the choice of the bandwidth. We show the results for three different bandwidths (see Appendix B for more information about bandwidth choice). As we have data for multiple years since 2005 we pool the data for the relevant years (and include year dummies) to improve the precision of our estimates. Finally, systematic differences in missing values between schools might bias the results. To probe the randomness of the missing values we have estimated our main regression models using a dummy for missing outcome variable as dependent variable. These analyses don't yield concerns about systematic differences in missing outcomes which might bias our results.

5. The effect of the CSCC-program on the school choice of teachers

The CSCC-program explicitly aimed at getting a more experienced and better qualified teaching staff in the targeted schools in the poor areas. As a first step in the analysis we investigate whether the school choice of teachers is affected by the difference in teacher pay induced by the program. Did the program succeed in getting more experienced teachers in the targeted schools in the poor areas?

A first impression of the effect of the program can be obtained from Figures 1 and 2. The top panel of Figure 1 shows the relationship between the average years of experience of the teaching staff in schools and the poverty score that was used for the eligibility for the CSCC-program. This relationship is shown for 2005, the start of the program, for 2010, the last year when there was full-compliance with the program, and for 2013, the last year for which we have data. The bottom panel shows the relationships with the tenure of the teaching staff measured in years. Schools with poverty scores above zero were eligible for the program.

In the left figure of the top panel we observe a downward sloping relationship between the experience of the teaching staff and the poverty score. Schools in areas with a lower poverty score have a more experienced teaching staff. This relationship directly reflects the preferences of more experienced teachers as these teachers may first select a school from the available teaching slots. In 2005, at the start of the program, there is no difference in the experience of the teaching staff of schools at both sides of the cut-off. In 2010, however, the levels of experience at the cut-off are very different. Schools at the right of the cut-off, those that were eligible for the program, have a more experienced teaching staff than schools at the left of the cut-off. In 2013 the difference in experience was reduced. This can be explained by a reduction of program eligibility since 2010.

Changes in the tenure of the teaching staff are shown in the bottom panel of Figure 1. In 2005 we also observe a downward slope for tenure suggesting that teachers not only prefer to work in schools in richer areas but also prefer to work more years in these schools. At the cut-off tenure is slightly higher at the left side. In 2010 the slope has become flatter and at the cut-off tenure is now slightly higher on the right side. In 2013 this difference appears to be slightly larger than in 2010.

Figure 2 further illustrates what happened over time. As we have data about experience and tenure since 2002 we can include several additional pre-treatment years. The figure shows the differences in program participation, experience and tenure for schools at the cut-off for program eligibility (the Y-axis) for each year between 2002 and 2013. These differences are obtained as point estimates from regression discontinuity models based on Equation (1) (see Table A.6 in the appendix for all the estimation results). The top panel of Figure 2 shows that the differences in program participation were small between 2002 and 2007[15]. Program participation increased strongly for the eligible schools in 2008, and full compliance was achieved in 2009 and 2010. After 2010 the difference in program participation declined fast which can be explained by the change in the eligibility rules in 2011. The figure also shows that the steepest increases and decreases in program participation difference took place in the smallest discontinuity samples. Hence, for schools close to the cut-off the largest differences in program participation occurred between 2008 and 2011. The middle panel shows the differences in experience of the teaching staff at the cut-off. We observe that these differences were very small between 2002 and 2008. In the next years we observe a sharp increase in the experience difference. Schools that were eligible for the program obtain a more experienced teaching staff than schools that were not eligible. The difference in experience increases to two to three years between 2009 and 2011, and decreases gradually in the next years. The largest differences are found for schools in the smallest discontinuity sample of schools across the cutoff. The bottom panel shows the changes in the tenure of the teaching staff. In the first years the differences in tenure are less constant than the differences in experience and fluctuate around zero. However, since 2008 we also observe that the differences in tenure start to increase. In 2010 and 2011 the difference is approximately one year. The increases in tenure are smaller than the changes in experience.

These figures show that the changes in program participation precede and are consistent with the changes in experience and tenure across the cut-off. Changes in program participation are steeper than changes in experience and tenure. Apparently, it takes some time before teachers adjust their school choices upon changes in the program. Figure 2 also shows that the period

[15] The estimate for the smallest discontinuity sample in 2007 suggests that the 'wrong schools' were treated in this year. However, in 2007 non-eligible schools were not treated. The negative estimate can be explained by a special pattern of treated and untreated schools at the right side of the cutoff. The untreated schools were schools close to the cut-off. If we exclude the forcing variable from the model we find a positive point estimate of 0.37 (0.05).

since 2009 is the most interesting period for our analysis. In these years schools that were eligible for the program have obtained a more experienced teaching staff than schools that were not eligible for the program, and this difference might translate in an improvement of the student outcomes. In the earlier years there was not yet a re-allocation of experienced teachers across the cut-off. We, therefore, focus our main analysis of the impact of the program on teacher outcomes and student outcomes on the years since 2009.

Next, we proceed by estimating the main models from Section 4. We start by estimating Equations (1) to (3) using a 1st-order and 2nd-order specification of the forcing variable, which is allowed to be different at either side of the cutoff. The models also control for grade level, year and a quadratic for school size. Columns (1) to (3) of Table 2 show the effects on teacher experience and columns (4) to (6) show the effects on tenure. The top panel shows the reduced form estimates, the middle panel shows the first stage estimates and the bottom panel shows the IV-estimates. The estimates are shown for three discontinuity samples around the cut-off for program eligibility. To improve the precision of the estimates we have pooled the data over the years since 2009. For each year we use data at the grade level, hence the data are at the grade X school X year level. We correct the standard errors for clustering at the school X year level, which is the level of the treatment[16].

The estimates in Table 2 show that the CSCC-program increased the experience of the teaching staff in the targeted schools. Program eligibility increased the average experience of the teaching staff by two to three years. These estimates are based on a pooled sample of five years and can therefore be seen as an average effect over these five years. The largest effects are found for the smallest discontinuity samples. This probably reflects the differential effect of the program for the two sides of the discontinuity. Schools on the left side of the cut-off lose experience whereas schools on the right side of the cut-off gain experience (see also Figure A.4 in the appendix). As mentioned earlier, the estimated effects should be interpreted as the combined effect of the changes at both sides of the cut-off. The IV-estimates show that the effects of actual participation in the program are even larger and vary between three and seven years. For the smallest discontinuity sample we even find a larger effect but the first stage

[16] Schools are treated every year, like in Leuven et al. (2007).

estimate might suffer from a weak instrument problem (the F-value of the excluded instrument is only 5.9). For the other estimates there are no concerns about a weak instrument problem. These IV-estimates should be interpreted as local average treatment effects. However, in 2009 and 2010 there was full compliance with the eligibility rules. This means that for these years the estimates apply to all eligible schools.

The right panel of Table 2 shows the estimates of the effect of the CSCC-program on tenure. The sample sizes are slightly different from the left panel because of missing values on teacher experience or tenure. The estimates show that the program also increased tenure with approximately one year. Moreover, the IV-estimates show that the effect of actual participation in the program is larger, and varies between one and two years. Again we find that the estimated effects increase when the estimation samples move closer to the cut-off. In addition, all the estimated effects are statistically significant. The estimated effects on tenure are smaller than the estimated effects on experience.

The policy could also induce a higher turnover of teachers. We have investigated this by looking at teachers that were new at a school (zero years of tenure) using the individual teacher data in our database.[17] We used this variable as a measure for turnover and analysed the effect of the program on the probability that a teacher is new at a school. We performed this analysis for each specific year and for the period 2009-2013, which is the main period of our analysis. We find that there is more turnover in eligible schools in the years of the implementation of the reform (2007 and 2008). However, our main results on experience and tenure concerns the period 2009-2013. For these years we don't find a difference in turnover; all estimates are small and statistically insignificant. This result also implies that the increase in tenure that we have found for this period is not driven by a decrease in turnover in these years. The increase in tenure results from a different type of turnover. Teachers with more tenure are less likely to leave program schools than to leave non-program schools.

[17] We cannot directly observe turnover because our database doesn't have individual teacher identifiers.

Robustness

The results in Table 2 are robust to different specifications of the forcing variable, such as a cubic or polynomial specification (Calonico et. al. 2014), or to including additional controls (Table A.5 in the appendix shows the estimation results). In addition, the estimated effects are quite similar if we use school-level data instead of grade-level data. We have also estimated the effect of program eligibility for each specific year since 2005 (see Table A.6 in the appendix). Excluding the last years from the estimation sample slightly increases the estimated effects. Including data from 2008 in the estimation sample reduces the estimated effects for experience to 1.2 to 1.4 years and for tenure to 0.3 to 0.6. This reduction of the estimated effects directly follows from the timing of the implementation of the program and the zero-effects in 2008, as shown in Figure 2. These sensitivity analyses suggest that the results in Table 2 are probably conservative estimates of the effect of the program on the school choice of teachers.

Finally, we have also estimated models that include school fixed effects using data for the whole period 2005-2013. We also include the first years after the reform of program to fully exploit the within school variation in program participation from the implementation years of the reform (see Figure A.1). We find that program participation on average increases experience with 1.7 to 1.9 years and tenure with 0.4 to 0.5 years. The fixed effect estimates are smaller than the IV-estimates from our RD-models. It should be noted that the fixed effects estimates exploit variation within schools over time and treat the variation on both sides of the cutoff similarly, whereas the RD-estimates pick up what happens at both sides of the cut-off.

In sum, we find that eligibility for and participation in the CSCC-program induced a re-allocation of experienced teachers towards the targeted schools between 2009 and 2013. The estimated effects on teacher experience are large and imply that the school choice of teachers is responsive to differences in teacher pay. The program also increases tenure in the targeted schools but these effects are smaller than the effects on teacher experience. As experience consists of tenure plus experience obtained at other schools this implies that the effect on teacher experience is mainly driven by the hiring of teachers from other schools. Approximately 20 % of teachers is new at a school each year; on average 2 to 3 teachers per school. The estimated effects imply that each new teacher brings approximately 6 to 12 years additional experience to an eligible school.

6. The effect of the CSCC-program on the performance of students

The second and most important aim of the program was to improve the educational outcomes of students in the disadvantaged areas. In this section we investigate the effect of the program on student performance using five measures: insufficient attendance, grade retention, dropout and test scores in math and language. We start by investigating the effect on three non-test scores measures which are available for all grades and all schools in Uruguay. Next, we analyze the effect on test scores that are obtained from samples of schools.

The effect of the CSCC-program on non-test-score outcomes

Figure 3 gives a first impression of the effect of the program on the three non-test-score measures. The relationship between these three measures of student performance and the poverty score of the schools is shown for 2005, 2010 and 2013. For all three outcomes we observe an upward sloping trend indicating that schools with higher scores on the poverty index have worse student outcomes. Insufficient attendance, grade retention and dropout are more likely in schools in poor areas. Most importantly, the figures for 2005 are remarkably similar to those in 2010 and 2013. There appear to be no major changes in the outcomes for schools at the cut-off. This suggests that the policy didn't have a large effect on student performance in the targeted schools.

Figure A.5 shows changes in program participation and student outcomes since 1992, which is the complete period for which we have data on these outcomes. The Y-axis shows the differences between schools at the two sides of the cut-off. To improve comparability, we have standardized the student outcome variables with mean zero and standard deviation of one. The top panel, which is similar to the top panel in Figure 2, further illustrates that over a long period before 2008 there were no differences in program participation for schools at the cut-off. The three figures on student outcomes don't provide a clear and consistent picture about the impact of the program. Insufficient attendance seems to decline over time, grade retention seems to increase over time and drop out seems to be fairly constant. The changes in insufficient attendance overlap with the years in which program participation increased. However, the decrease in insufficient attendance in 2007 precedes the increase in program participation with one year, and precedes the increase in teacher experience with two years. Moreover, the reduction in insufficient attendance is sensitive to the discontinuity sample that is used for the estimation. Hence, the figure suggest that the program might have had some impact on insufficient attendance but probably had no impact on grade retention or drop out.

Next, we have estimated the main models of Section 4. Table 3 shows the estimates of the effect of the program using reduced form and IV-models as specified in Equations (1) to (3). The models include a 1^{st}-order or 2^{nd}-order specification of the forcing variable. As in Table 2, we have pooled the data over the years since 2009. Columns (1) to (3) show the effects on insufficient attendance, columns (4) to (6) show the effects on grade retention and columns (7) to (9) show the effects on drop out. The estimation samples are approximately ten percent larger than the samples for the teacher outcomes in Table 2 due to missing values on experience or tenure.

The estimates of the effect of the CSCC-program on student outcomes are less clear than the estimates of the effect on teacher outcomes[18]. We don't find evidence that the program reduced grade retention or drop out; all estimated effects of (eligibility for) the CSCC-program on grade retention and drop out are statistically insignificant or have the wrong sign. The program appears to have some impact on insufficient attendance. We observe that most point estimates are negative and some estimates are statistically significant. However, the results are sensitive to changes in the specification of the forcing variable. In addition, the effects are not consistent across the discontinuity samples and don't increase when the discontinuity samples get smaller, as they did with teacher outcomes. We have also estimated the effect of program eligibility for each specific year since 2005 (see Tables A.7 to A.9 in the appendix). These estimates do not reveal a clear pattern about the impact of the program on student outcomes.

We have investigated the robustness of these estimates by performing the same set of sensitivity analyses as with the teacher outcomes in the previous section (see Table A.10 in the

[18] It should be noted that the results at the aggregate grade level include any teacher peer effects (Jackson & Bruegmann 2009).

appendix). In general, the estimated effects of the program on student outcomes in these robustness analyses are somewhat smaller than the results shown in Tables 3 and 4.

We further investigated the effectiveness of the program by looking at differences between grades and by constructing a more detailed attendance variable. The effects of the program might differ among grades as previous studies have found that teacher experience is especially important in early grades (Krueger, 1999; Chetty et al., 2011; Gerritsen et al., 2017) A further advantage of looking specifically at earlier grades is that there is more variation in the outcome variables. However, the estimates for early and later grades also don't provide clear support for the effectiveness of the program (see Table A.11). The program seems to reduce grade retention in the early grades but also seems to increase grade retention in the other grades. Moreover, the program seems to reduce drop out in grades three to six but this result is sensitive to the specification and discontinuity sample. The effects on insufficient attendance are found through all grades but the estimated effects are larger for the early grades, which is consistent with previous findings about the effectiveness of experienced teachers. Again these results are sensitive to the specification and restrictions about the discontinuity sample. The estimated effects for the higher grades are smaller. [19]

To improve the precision of our estimates we have constructed an additional outcome variable about school attendance. This variable is based on the two previous measures 'insufficient attendance' and 'dropout', and on additional information about the number of days that students attended school (see Section 3). The estimates of the effect on this new attendance measure confirm the previous findings (Table A.12). The CSCC-program appears to increase student attendance, especially for students in early grades, but the results are sensitive to the specification of the forcing variable and restrictions on the discontinuity samples.

Furthermore, we have investigated the impact of years in the program. It might be possible that the impact of the program on students will not be immediate, or might depend on the intensity of the treatment (number of years in the treated state). The results based on IV-models in which the number of treatment years is instrumented with program eligibility are

[19] The data also allow us to separately investigate the effects of the program on male and female grade retention. We find that the effects for boys and girls are similar to those reported in Table 4.

consistent with the previous IV-estimates, and don't provide clear evidence about the effectiveness of the program.

The effect of the CSCC-program on Test Scores

Our second data source, the National Evaluation of Learning project, provides data on test scores that have been collected in representative samples of schools in 1999, 2002 and 2009. Students in grade 6 had to take standardized tests in math and language (see Section 3). To fully exploit these data we use two features in our analysis. First, we use the variation induced by the eligibility rules of the CSCC-program in 2005. We investigate the effect of the CSCC-program on the cognitive achievement scores by estimating the reduced form model as specified in Equation (1). As there was full compliance in 2009 the reduced form estimates are equal to the IV-estimates. Second, we use the time dimension of our data within a difference-in-differences framework[20]. This allows us to observe whether the relative performance of students at the right side of the cut-off has changed over time. Combining the RD and the difference-in-differences framework requires the additional assumption that the RD-estimate does not differ before and after the introduction of the policy.[21]

Table 4 shows the RD-estimates for math and language test scores in 2009, and the results of the difference-in-differences approach. Again we use discontinuity samples of schools across the cutoff. The estimation samples are smaller than in the previous analysis as test scores are only available for a sample of schools and not for the whole population. The rows show the effects of being eligible for the program in 2005 on test scores using different specifications of the forcing variable. Test scores have been standardized with mean zero and standard deviation of one.

[20] The second difference is whether the tests were taken before or after the redesigning of the CSCC-program in 2005.

[21] Jackson (2017) points out that this so-called DiRD assumption may be violated if, for instance, the quality of students or the quality of schools at the cut-off changes over time.

We start by looking at the results for 2009. In this year all schools in our sample at the right side of the cut-off and no schools at the left side of the cut-off participated in the program. The estimates of the effect of the program on Math test scores are all statistically insignificant and nearly all point estimates are negative. The results for the language test are quite similar and also don't yield evidence for a positive effect of the CSCC-program. We also estimated these models using test scores from 1999 and 2002 (see Table A.13 in the appendix). The difference in test scores across the cut-off in 1999 and 2002 appears to be different from the difference we observed in 2009. All point estimates are positive and we also find one statistically significant effect, indicating that students at the right side of the cut-off performed better in these years than schools at the left side of the cut-off. The difference-in-differences estimates summarize the changes in relative performance of students in schools at the right side of the cut-off over time. We find that all point estimates are negative and some point estimates have p-values close to the regular significance levels. Hence, these estimates don't provide evidence that the relative performance of students in schools at the right side of the cut-off has improved.

In sum, we have investigated the effect of the program on five measures of student performance. These estimates don't provide a compelling case for the effectiveness of the program on student outcomes. Based on these results, and even without adjustment for multiple testing, we cannot reject the null hypothesis that the program had no effect on student outcomes. The consistency of these results across a range of non-test-score measures and test-score outcomes leads us to conclude that the impact of the CSCC-program on student performance is likely to be quite small.

7. Mechanisms: why are the effects on student performance modest?

We have found that the CSCC-program has a large effect on experience of the teaching staff of schools in poor neighborhoods and also increases tenure in these schools. However, we have also found that the effects on student performance appear to be modest. This raises the question why the increase of experience did not have a substantial effect on student performance.

Does the policy affect the relevant parts of the experience distribution?

The modest effect of the program might come from the changes in the experience distribution. In a recent study Hendricks (2014) summarizes what is known about the relationship between teacher experience and student achievement. ‘Teacher performance improves dramatically in the first four years of teaching and then levels off in subsequent years.’ This implies that the impact of the CSCC-program on student performance not only depends on the increase in average experience of the teaching staff in the targeted schools but also on the changes in the experience distribution within schools that are affected. The program would be most effective if it especially replaces new teachers with experienced teachers (4+ years of experience). However, if the program replaces experienced teachers with even more experienced teachers we would expect a much smaller impact on student performance.

The individual teacher data in our database allow us to investigate the changes in the experience distributions of teachers in schools. Figure 4 shows the density functions of experience in schools in the smallest discontinuity sample around the cut-off in 2005 and in 2010. In 2005 the density functions for the two samples of schools around the cut-off are quite similar but schools that are eligible for the program have slightly more teachers with less than ten years of experience and have slightly less teachers with more than ten years of experience. In 2010 this pattern has reversed. Schools that are eligible for the program have less teachers with less than ten years of experience and have more teachers with more than 10 years of experience. This means that the program has affected the whole distribution of experience including both the early stages of the career as the later stages of the career.

To further assess the importance of the changes in the experience distribution we have estimated the effect of the program on ‘the most relevant parts of this distribution’. We constructed dummies for specific parts of the experience distribution; 0-4, 5-9, 10-14, 15-19, 20-24, 25-29, 30-34, 35-40 years of experience. Next, we estimated linear probability models of the effect of the program on these specific parts of the experience distribution. Table 5 shows the estimation results. Each row shows the effect of a separate regression of a dummy for an experience category on eligibility for the program and a set of controls. The left panel of Table 5 shows the estimates for 2006[22], the right panel shows the estimates for the period 2009-2013.

[22] We don’t use the data of 2005 because of a difference in the coding of the experience categories. This changed in 2006.

In the left panel of Table 5 we observe that there were only small differences in the experience categories between schools across the cut-off. In the period 2009-2013 we observe a different pattern. Teachers with no more than 10 years of experience are less likely to work at schools that were eligible for the CSCC-program. On the other hand, teachers with 11 to 20 years of experience are more likely to work in schools at the right side of the cut-off. The other experience categories remained largely unchanged. This means that the program induced replacements of teachers with no more than 10 years of experience by teachers with 11 to 20 years of experience. This pattern of replacements seems only partially efficient considering the recent literature on the productivity-experience profile of teachers.

We further examined this hypothesis in by looking at schools that had many or little 'rookie teachers' (less than 5 years of experience) at the start of the reform. We expect that schools that had many rookie teachers will be most likely to see reductions in rookie teachers. For this analysis we have split the sample of schools in two parts: schools that had an above median proportion of rookie teachers in 2005, and schools that had a below median proportion of rookie teachers in 2005.

Figure 5 gives a first impression of the impact of the program for schools with many rookie teachers (left panel) and for schools with little rookie teachers (right panel) in the largest discontinuity sample around the cut-off. The figures suggest that the impact of the program depends on the proportion of rookie teachers at the start of the reform. For schools with little rookie teachers we don't observe a difference in the outcomes of schools across the cutoff. However, the program appears to improve the results for schools with many rookie teachers at the start of the program. At the cutoff we observe less insufficient attendance, grade retention and school dropout for schools that were eligible for the program.

Next, we re-estimated the reduced form models from Table 3 for both samples. Table 6 shows the results for all grades and for the early grades only. We find a clear difference between the two samples of schools. The program improved student outcomes in schools that had many rookie teachers at the start of the reform. For this sample we find that the program reduced 'insufficient attendance' and 'drop out' for all students, and especially for students in grades 1 and 2. Moreover, the program also appears to have a beneficial effect on grade retention. For the sample of schools that had little rookie teachers in 2005 we don't find beneficial effects on student outcomes. We find similar results when we split the sample in schools that had many or little teachers with less than four years of tenure in 2005 (Table A.14). These results suggest that the program can be effective if it focuses on replacing very inexperienced teachers.

Hiring or keeping teachers?

A second argument for explaining the modest effect of the program might come from the distinction between experience obtained at the current school (tenure) and experience obtained at other schools. Since the seminal studies by Abraham & Farber (1987), Altonji & Shakotko (1987) and Topel (1991) it is well established that the wage returns to tenure exceed the wage returns to experience obtained in other firms. This difference in returns might result from firm specific human capital, but also from unobserved factors related to worker quality, job quality or match quality. The CSCC-program can have a different effect on these two components of experience. If the program is especially important for the experience component that is less important for student outcomes this would explain the modest effect. For instance, it has been found that specific experience of teachers has a larger impact on test score improvement than general experience of teachers (Jackson 2013; Ost 2014[23]).

For investigating the difference between 'hiring' and 'keeping' teachers we have decomposed 'total experience' in 'years of tenure' and 'years of experience at other schools' (total experience= tenure + experience obtained at other schools)[24]. Reduced form estimates of the effect of the CSCC-program show that the program affected both components of experience. However, the impact on experience obtained at other schools is larger, especially in the two discontinuity samples closest to the cut-off for program eligibility (see Table A.15).

Next, we have investigated the relationship between the two experience components and student outcomes. Tables A.16 and A.17 show estimates from regressions of student outcomes on 'total experience' or on the two components of 'total experience' (tenure or experience obtained at other schools). Table A.16 shows the estimates using non-test-score measures as outcomes, Table A.17 shows the estimates using test scores as outcomes.

[23] This study looks at grade specific experience and also at the timing of the experience components (recent or distant).

[24] Abraham & Farber (1987) apply the same decomposition.

The estimates in Table A.16, and especially those in Table A.17, reveal a remarkable pattern. First, experience has a small, and mostly statistical insignificant, association with (better) student outcomes. Second, tenure appears to be much more important, especially for insufficient attendance and for test scores in math and language. More years of tenure decrease insufficient attendance and improve test scores in math and language. Also note that these results are robust to including grade by school fixed effects (Table A.16). Third, increases in experience obtained at other schools don't improve student outcomes or even decrease student performance. Hence, the estimates for the two underlying components of experience are very different, especially for test scores. We also find that nearly all tests on the similarity of the estimates for the two experience components reject the null hypothesis that the estimates are equal. Hence, these estimates suggest that tenure is important for student performance and that experience obtained at other schools doesn't contribute or is even detrimental to student performance. It should be noted that these results are based on associations and might not reflect causal relationships between the experience components and student performance.

The consistency of the results across different outcomes and discontinuity samples is remarkable, and suggests that increases in tenure and increases in 'experience obtained at other schools' have a different effect on student performance. Hence, keeping teachers appears to be more beneficial for students than hiring teachers. These findings suggests that the CSCC-program especially has an impact on the experience component that appears not to be important for student performance.

8. Conclusion and discussion

This study investigates the effects of a policy program aimed at attracting more experienced and better qualified teachers in primary schools in poor neighborhoods in Uruguay. Teachers could earn substantially higher salaries by working in these schools. We find that the program was successful in increasing experience of teachers in schools in the targeted neighborhoods. For the period 2009-2013 the policy on average increased the experience of the teaching staff with two to three years and tenure with one year. Estimates of the treatment on the treated effect indicate effects of three to seven years of experience and one to two year of tenure.

Despite this substantial change in the composition of the teaching staff the impact on student performance appears to be small. We have investigated the impact of the program on a range of non-test-score outcomes and on test scores. Even without adjustment for multiple testing it is difficult to reject the null hypothesis that the program doesn't influence student outcomes. We find some evidence that the program increased the attendance of students, especially the attendance of students in grades 1 and 2. However, these results are not robust. We don't find an effect of the program on grade retention or drop out, nor on test scores in math and language. This leads us to conclude that the effects of the CSCC-program on student outcomes are likely to be quite small.

The modest results might be explained by two factors. First, the program induces replacement of teachers who are on the steepest part of the productivity-experience profile, but also replaces teachers who are on parts of the productivity-profile that are less steep. The latter replacements are expected to be less efficient, and thereby probably reduce the impact of the program. Consistent with this, we find a clear difference between schools that had many rookie teachers and schools that had little rookie teachers at the start of the reform. The program improved student outcomes in schools with many rookie teachers but didn't improve outcomes in schools with little rookie teachers. Second, the CSCC-program especially affects the hiring of experienced teachers from other schools. This appears to be the component of experience that is less relevant for student performance. Second,

This paper shows that the school choice of teachers is sensitive to variation in teacher pay. Schools in disadvantaged areas can attract or keep more experienced teachers by offering higher salaries. However, the impact on student performance appears to be small, probably because the program increased experience in ways that are not strongly associated with improved student outcomes. A redesign of the program towards teacher characteristics that actually improve student outcomes is expected to improve the results of the program. For instance, the program would probably become more effective if it would focus on: 'keeping teachers' instead of 'hiring teachers', replacing teachers with less than five years of experience and prioritizing early grades in primary education. Such a program is expected to improve the outcomes of students in poor areas. However, the reallocation of 'good teachers' will probably also harm the outcomes of students in areas that lose these teachers. Total welfare might increase if the gains in achievement for students in poor areas will also reduce potential negative spillovers from poor areas to other areas, by improving employment opportunities later in life and reducing socially harmful behavior such as criminal activities.

References

Aaronson, D., Barrow, L., Sander, W., 2007. Teachers and student achievement in Chicago public high schools. *Journal of Labor Economics*, 25, pp.95-135.

Abraham, K.G., H.S. Farber, 1987, Job Duration, Seniority, and Earnings. American Economic Review, 77 (3), pp. 278-297

Altonji, J.G. and R. A. Shakotko, 1987. Do Wages Rise with Job Seniority? Review of Economic Studies, 54 (3), pp. 437-459

Borghans, L., B.T. ter Weel, and B.A. Weinberg, 2008. Interpersonal styles and labor market outcomes. *Journal of Human Resources*, 43 (4), pp. 815-58.

Britton, J. and Propper, C., 2016. Teacher pay and school productivity: Exploiting wage regulation. *Journal of Public Economics*, 133, pp.75–89

Calonico, S., Cattaneo, M. and R. Titiunik, 2014. Robust Nonparametric Confidence Intervals for Regression-Discontinuity Designs. *Econometrica*, 82(6), pp.2295–2326.

Cattaneo, M., M. Jansson and X. Ma, 2017. Rddensity: Manipulation Testing based on Density Discontinuity, working paper.

Cattaneo, M.D., N.Idrobo, R. Titiunik, 2018. A practical introduction to Regression Discontinuity Designs: Volume I, Cambrigde University Press, forthcoming.

Chay, K.Y., P.J. McEwan and M. Urquiola, 2005. The Central Role of Noise in Evaluating Interventions That Use Test Scores to Rank Schools. *American Economic Review,* 95(4), pp.1237-1258.

Chetty, R., J.N Friedman, N. Hilger, E. Saez, D. Whitmore Schanzenbach and D. Yagan, 2011. How does you kindergarten classroom affect your earnings? Evidence from project STAR. *Quarterly Journal of Economics,* 126 (4), pp.1593-1660.

Chetty, R., J.N. Friedman and J.E. Rockoff, 2014. Measuring the Impacts of Teachers II: Teacher Value-Added and Student Outcomes in Adulthood. *American Economic Review*, 104(9), pp.2633-2678.

Clotfelter, C.T., H.F. Ladd and J.L. Vigdor, 2007. Teacher credentials and student achievement: longitudinal analysis with student fixed effects. *Economics of Education Review*, 26 (6), pp. 673–682.

Clotfelter, Ch., E. Glennie, H. Ladd and J. Vigdor, 2008. Would higher salaries keep teachers in high-poverty schools? Evidence from a policy intervention in North Carolina. *Journal of Public Economics*, 92 (5–6), pp.1352–1370.

Clotfelter, Ch., H. Ladd and J. Vigdor, 2011. Teacher mobility, school segregation, and pay-based policies to level the playing field. *Educational Finance Policy,* 6 (3), pp.399–438.

De Ree, J., K. Muralidharan, M. Pradhna and H. Rogers, 2018. Double for Nothing? Experimental Evidence on the Impact of an Unconditional Teacher Salary Increase on Student Performance in Indonesia, Quarterly Journal of Economics, 133 (2), pp. 993-1039.

Dolton, P. and O.D. Marcenaro-Gutierrez, 2011. If you pay peanuts do you get monkeys? A cross-country analysis of teacher pay and pupil performance. *Economic Policy,* 26 (65), pp.5–55.

Dolton, P. and W. van der Klaauw, 1995. Leaving teaching in the UK: a duration analysis. *Economic Journal*, 105 (429), pp.431–444.

Dolton, P. and W. van der Klaauw, 1999. The turnover of teachers: a competing risks explanation. *Review of Economics and Statistics*, 81 (3), pp.543–550.

Finan, F., B. A. Olken and R. Pande, 2015. The Personnel Economics of the State. *Handbook of Economic Field Experiments***,** Forthcoming.

Frederikson, P., B. Ockert and H. Oosterbeek, 2013. Long-term effect of class size. *Quarterly Journal of Economics*, 128 (1), pp.249-285.

Gerritsen, S., Plug, E. and D. Webbink, 2017. Teacher quality and student achievement: Evidence from a sample of Dutch twins. *Journal of Applied Econometrics, 32 (3), pp.643-660.*

Gershenson, S, 2016. Linking teacher quality, student attendance, and student achievement. Education Finance and Policy, 11(2), pp. 125-149.

Gilpin, G.A., 2011. Re-evaluating the effect of non-teaching wages on teacher attrition. *Economics of Education Review*, 30 (4), pp.598-616.

Hanushek, E.A, 1997. Assessing the effects of school resources on student performance: an update. *Educational Evaluation Policy Analysis*, 19 (2), pp. 141–161.

Hanushek, E.A, 2003. The failure of input-based schooling policies. *Economic Journal,* 113 (485), pp.F64–F98.

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

Hanushek, E.A. and J.R. Rivkin, 2010. Generalizations about Using Value-Added Measures of Teacher Quality. *American Economic Review: Papers & Proceedings,* 100 (May 2010), pp.267–271

Hanushek, E.A., 2011. The Economic Value of Higher Teacher Quality. *Economics of Education Review*, 30(2), pp.466-479.

Harris, D.N. and T.R. Sass, 2011. Teacher training, teacher quality and student achievement. *Journal of Public Economics*, 95, pp.798-812.

Heckman, J.J., R. Pinto and P. Savelyev, 2013. Understanding the mechanism through wich an influential early childhood program boosted adult outcomes. *American Economic Review.* 103(6), pp. 2052-86.

Heckman, J.J. and Y. Rubinstein, 2001. The importance of non-cognitive skills: Lessons from the GED testing program. *American Economic Review*, 91(2), pp. 145-49.

Heckman, J.J., J.Stixrud and S.Urzua, 2006. The effects of cognitive and noncognitive abilities on labor market outcomes and social behavior. *Journal of Labor Economics*, 24(3), pp. 411-82.

Hendricks, M., 2014. Does it pay to pay teachers more? Evidence from Texas. *Journal of Public Economics*, 109, pp.50-63.

Imazeki, J., 2005. Teacher salaries and teacher attrition. *Economics of Education Review*, 24 (4), pp.431–449.

Imbens, G. and J. Angrist (1994). Identification and Estimation of Local Average Treatment Effects, *Econometrica*, 62, 467-76.

Jackson, C.K and E. Bruegmann, 2009. Teaching students and teaching each other: The importance of peer learning for teachers, *American Economic Journal: Applied Economics,* 1(4), 85-108.

Jackson, C.K, 2009. Student Demographics, Teacher Sorting, and Teacher Quality: Evidence from the End of School Desegregation, *Journal of Labor Economics*, 27 (2), 213-256.

Jackson, C.K, 2013. Match Quality, Worker Productivity, and Worker Mobility: Direct evidence from Teachers, *Review of Economics and Statistics*, 95(4), 1096-1116.

Jackson, C.K, 2017. The effect of single-sex education on test scores, school completion, arrests and teen motherhood: Evidence from school transitions, *NBER Working Paper 22222*.

Jackson, C.K., 2018. What do test scores miss? The importance of teacher effects on non-test-score outcomes, *Journal of Political Economy*, 126 (5), 2072-2107.

Jackson, C.K., R.C. Johnson and C. Persico, 2016. The effects of school spending on educational and economic outcomes: evidence from school finance reforms, *Quarterly Journal of Economics, 157-218.*

Kane, T.J. and D.O. Staiger, 2008. Estimating teacher impacts on student achievement: an experimental evaluation, NBER working Paper 14607.

Krueger, A.B., 1999. Experimental estimates of education production functions. *The Quarterly Journal of Economics,* 114 (2), pp.497-532.

Ladd, H.F. and L.C. Sorenson, 2017. Returns to Teacher Experience: Student Achievement and Motivation in Middle School, *Education Finance and Policy*, 12(2), pp. 241–279.

Leuven, E., M. Lindahl, H. Oosterbeek and D. Webbink, 2007. The Effect of Extra Funding for Disadvantaged Pupils on Achievement. *Review of Economics and Statistics, 89(4), pp.721-736.*

Lindqvist, E. and R. Vestman, 2011. The labor market returns to cognitive and noncognitive ability: Evidence from the Swedish enlistment. *American Economic Journal: Applied Economics*, 3(1), pp. 101-28.

Loeb, S., Page, M., 2000. Examining the link between teacher wages and student outcomes: the importance of alternative labor market opportunities and nonpecuniary variation. *Review of Economics and Statistics*, 82 (3), pp. 393-408.

Manacorda, M., (2012). The cost of grade retention. *Review of Economics and Statistics*, 94 (2), pp.596-606.

Murnane, R.J. and R.J. Olsen, 1990. The effects of salaries and opportunity costs on length of stay in teaching: evidence from North Carolina. *Journal of Human Resources,* 25 (1), pp.106–124.

Ost, B., 2014. How do Teachers Improve? The Relative Importance of General and Specific Human Capital. *American Economic Journal: Applied Economics*, 6(2), pp. 127-151.

Papay, J.P. and M.A. Kraft, 2015. Productivity Returns to Experience in the Teacher Labor Market: Methodological Challenges and New Evidence on Long-Term Career Improvement. *Journal of Public Economics,* 130, pp.105-119.

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

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

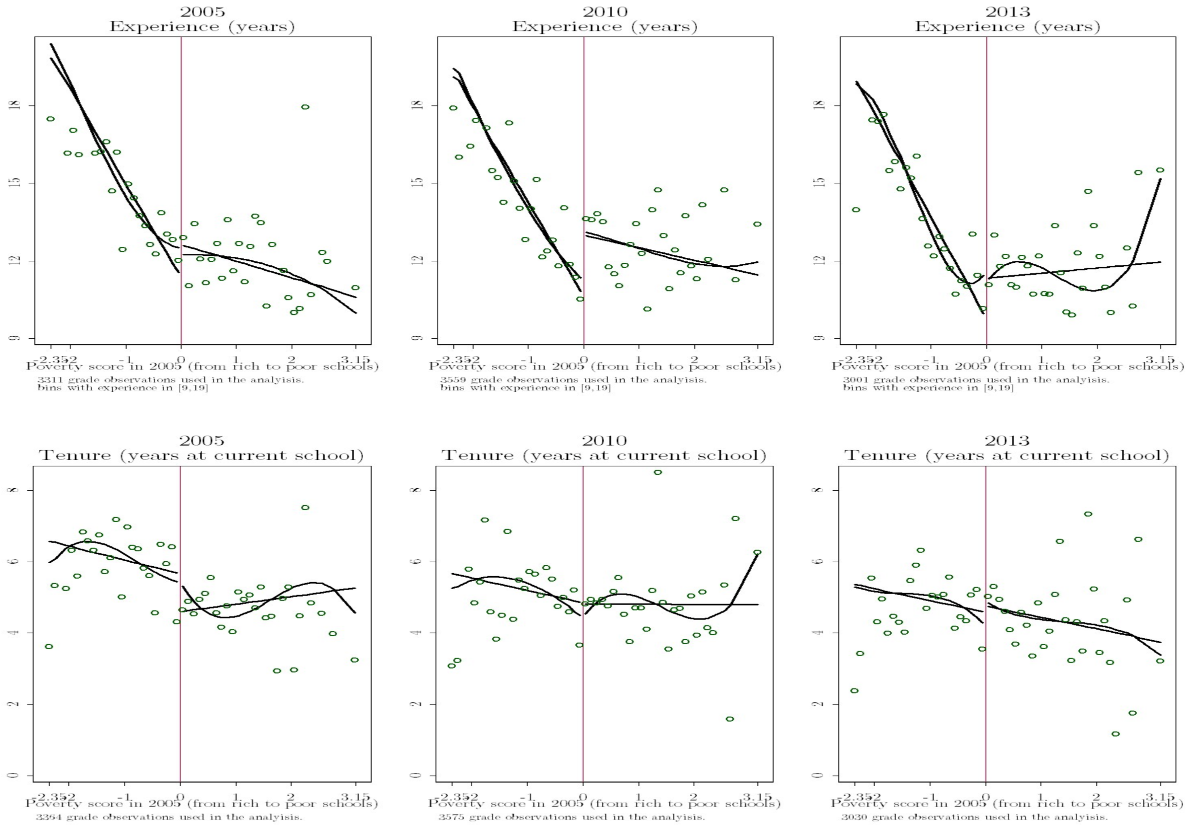
Topel, R., 1991. Specific Capital, Mobility, and Wages: Wages Rise with Job Seniority, Journal of Political Economy, 99 (1), pp. 145-176

Urguiola, M., 2006. Identifying class size effects in developing countries: Evidence from rural Bolivia. *Review of Economics and Statistics*, 88 (1), pp.171-177.

Wiswall, M., 2013. The dynamics of teacher quality. *Journal of Public Economics*, 100, pp.61-78.

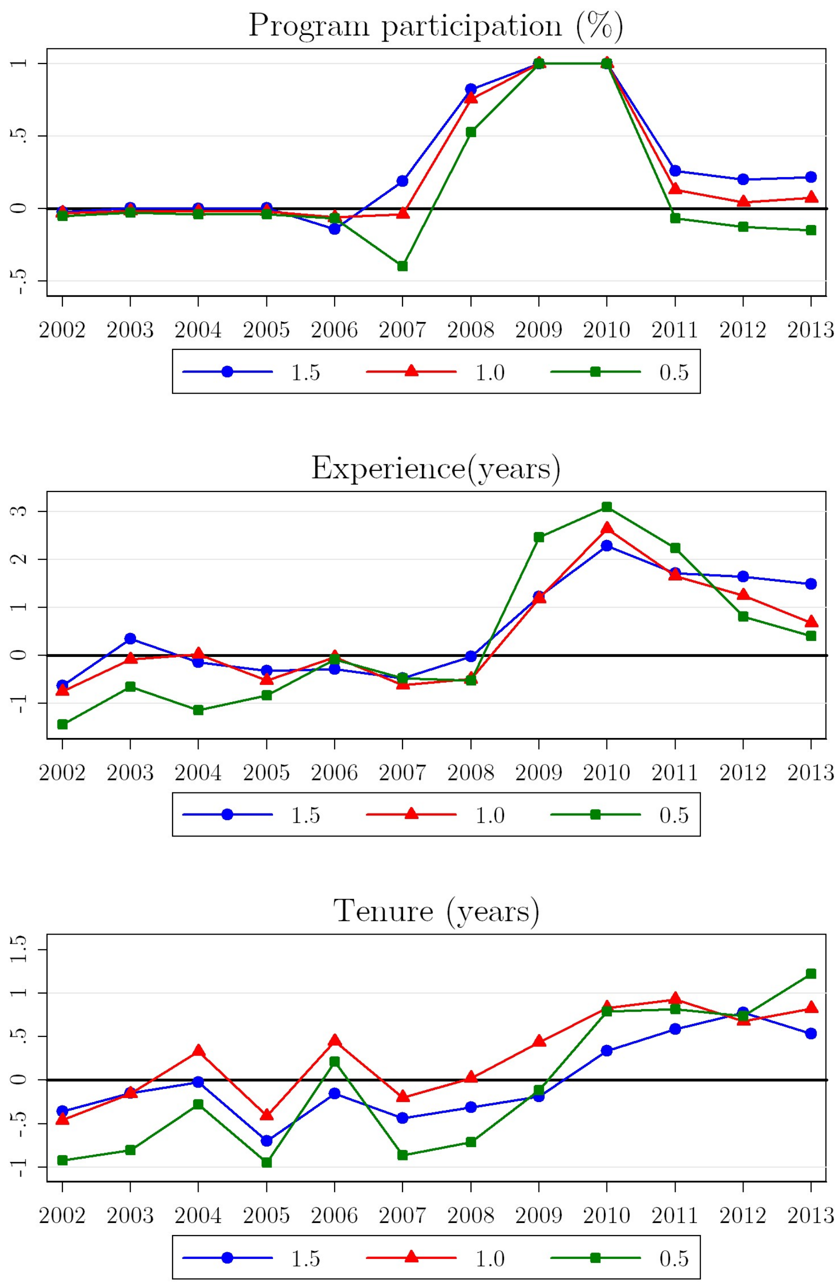
Figures

Figure 1. Teacher experience and tenure by poverty score in 2005, 2010 and 2013



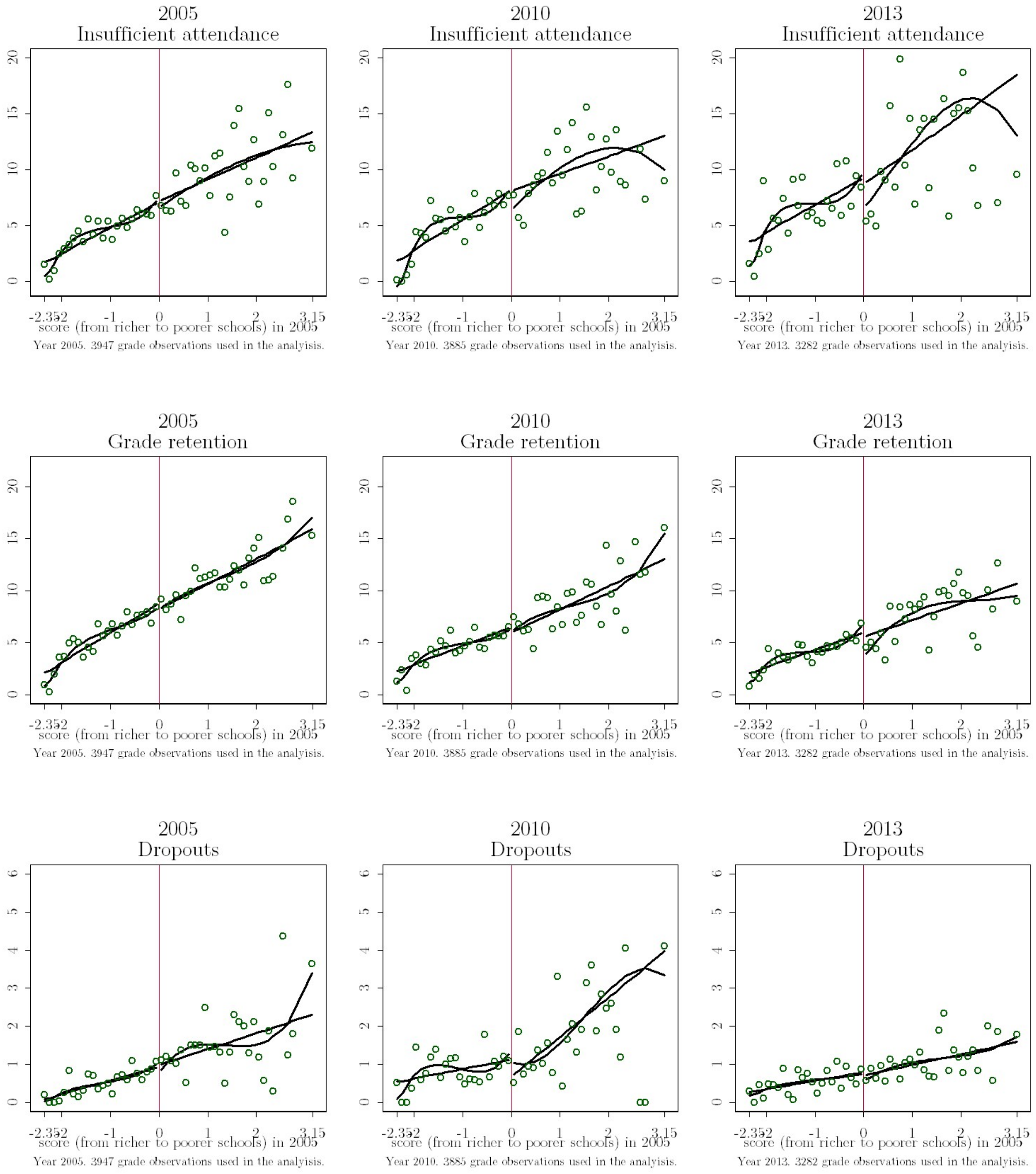
Notes: Each dot in the figure represents the mean of the dependent variable (experience or tenure) for schools located within a bin of width 0.1 of the poverty score. The figures use a linear and cubic fit for the regression lines at both sides of the cut-off for eligibility for the program.

Figure 2. Differences in program participation, experience and tenure for schools at the cut-off using three discontinuity samples 2002-2013



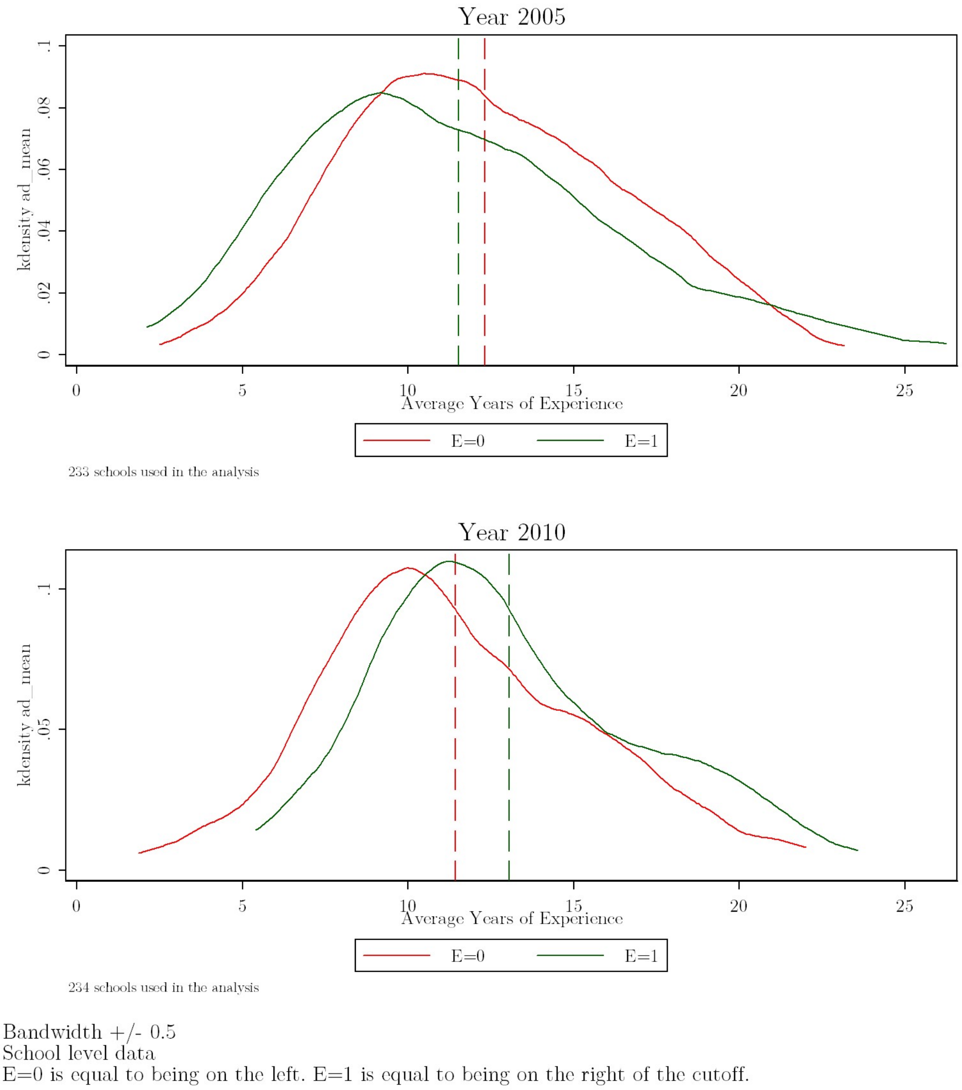
Notes: The figures show point estimates from regression-discontinuity models with a linear specification of the forcing variable and controlling for grade, year and a quadratic of school size. The three discontinuity samples used are +/- 1.5, +/- 1.0 and +/- 0.5 points of the running variable.

Figure 3. Student outcomes by poverty score in 2005, 2010 and 2013



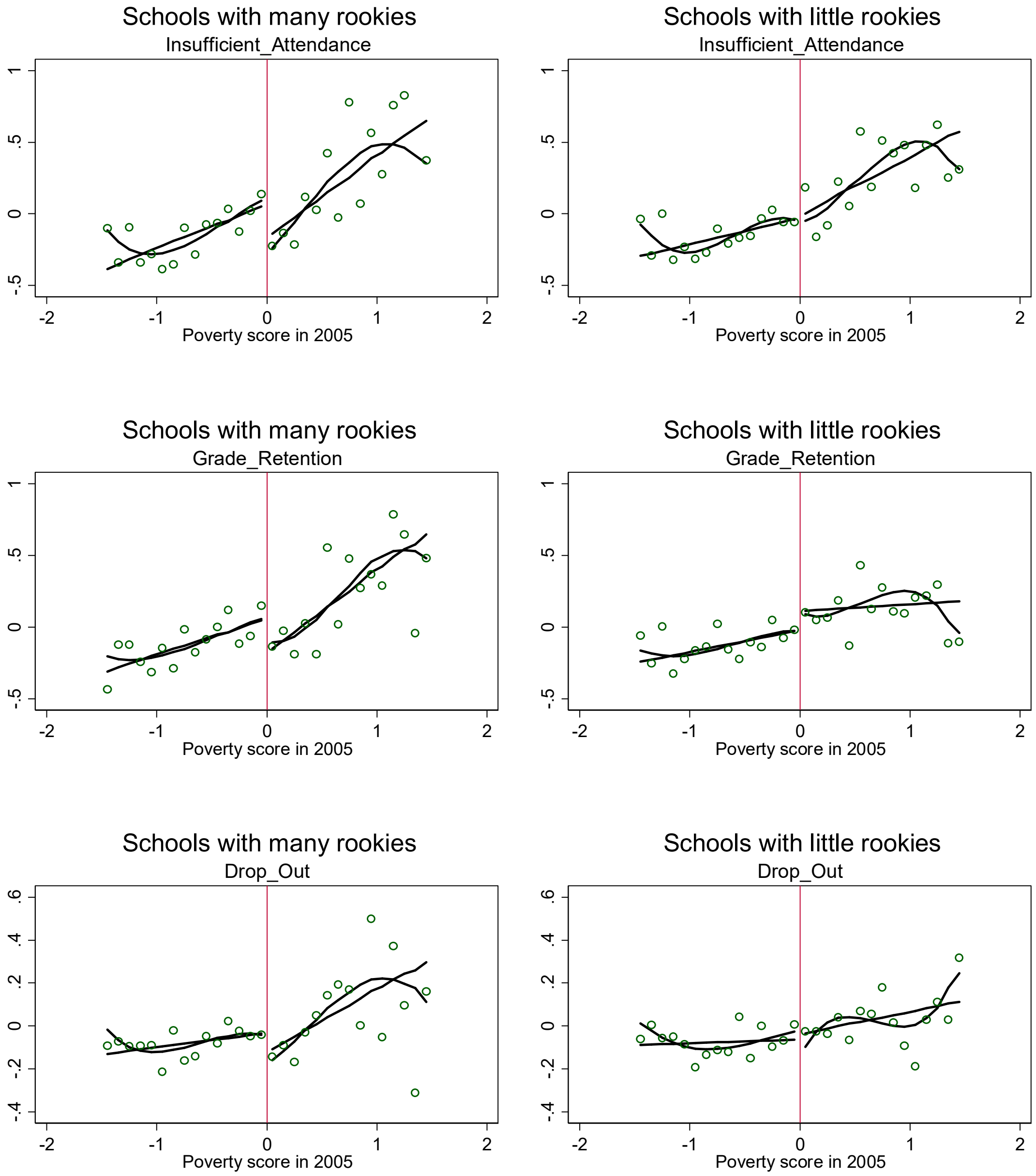
Notes: Each dot in the figure represents the mean of the dependent variable (insufficient attendance, grade retention or dropout) for schools located within a bin of width 0.1 of the poverty score. The figures use a linear and cubic fit for the regression lines at both sides of the cut-off for eligibility for the program.

Figure 4. Experience distributions in 2005 and in 2010 by program eligibility (individual teacher data)



Notes: The figures show density functions of experience in schools in the discontinuity sample of +/- 0.5 standard deviation of the poverty score around the cut-off in 2005 and in 2010. Both figures are based on individual teacher data.

Figure 5. Student outcomes by poverty score 2009-2013 for schools with many or little rookie teachers in 2005



Notes: Rookie teachers are teachers with less than 5 years of experience. School with many (little) rookie teachers have above (below) median proportion of rookie teachers in 2005. Each dot in the figure represents the mean of the dependent variable (insufficient attendance, grade retention or dropout) for schools located within a bin of width 0.1 of the poverty score. The figures use a linear and cubic fit for the regression lines at both sides of the cut-off for eligibility for the program.

Tables

Table 1 - Summary statistics of schools by program eligibility in 2005

	All schools					
	Not eligible		Eligible		Difference	
	(1)	(2)	(3)	(4)	(5)	(6)
	mean	st. dev.	mean	st. dev.	T-stat	sign
Panel A: Families and schools						
Poverty score	-0.79	(0.55)	0.85	(0.65)	34.15	***
Mothers with primary education or less	35.74	(15.87)	61.68	(12.63)	23.35	***
Children with unmet basic needs	28.20	(14.53)	58.08	(14.11)	26.56	***
Number of students	315	(162.43)	318	(187.80)	0.21	
Number of teachers	11.04	(4.92)	11.38	(5.85)	0.80	
Number of groups	11.32	(4.43)	11.67	(5.22)	0.92	
Panel B: Teacher outcomes						
Experience (in years)	14.30	(5.21)	12.20	(5.07)	-4.95	***
Tenure (in years at current school)	5.98	(2.85)	4.72	(2.59)	-5.66	***
Panel C: Student outcomes						
Insufficient Attendance	5.27	(3.48)	9.17	(6.13)	9.61	***
Grade Retention	6.58	(3.56)	10.41	(4.81)	11.25	***
Dropouts	0.65	(0.90)	1.41	(1.64)	7.11	***
School climate: Verbal violence	0.43	(0.50)	0.58	(0.49)	4.00	***
School climate: Physical violence	0.39	(0.49)	0.55	(0.50)	4.23	***
Panel E: Observations						
# Schools	374		284			
# Grades	2,295		1,752			

Notes: Standard deviations in parentheses. T-statistic of the difference in means. *** p<0.01, ** p<0.05, * p<0.1.

Table 2. Estimates of the effect of the CSCC program on experience and tenure 2009-2013

	Teacher experience				Tenure	
	(1)	(2)	(3)	(4)	(5)	(6)
Discontinuity sample	+/- 1.5	+/- 1.0	+/- 0.5	+/- 1.5	+/- 1.0	+/- 0.5
Reduced form						
1st-order polynomial	1.686***	1.526***	1.873***	0.403**	0.737***	0.657**
	(0.304)	(0.359)	(0.501)	(0.191)	(0.220)	(0.297)
2nd-order polynomial	1.703***	2.315***	3.344***	0.989***	0.640**	1.477***
	(0.449)	(0.546)	(0.726)	(0.261)	(0.314)	(0.387)
IV estimations						
First stage						
1st-order polynomial	0.573***	0.481***	0.364***	0.572***	0.482***	0.361***
	(0.029)	(0.037)	(0.059)	(0.029)	(0.037)	(0.058)
2nd-order polynomial	0.387***	0.315***	0.234**	0.388***	0.314***	0.233**
	(0.049)	(0.062)	(0.096)	(0.049)	(0.062)	(0.096)
Second stage						
1st-order polynomial	2.877***	3.213***	5.135***	0.692**	1.542***	1.806**
	(0.536)	(0.749)	(1.493)	(0.338)	(0.476)	(0.889)
2nd-order polynomial	4.320***	7.186***	13.038**	2.462***	1.993*	5.868**
	(1.165)	(2.048)	(5.153)	(0.741)	(1.089)	(2.821)
Observations	13,749	10,341	5,868	13,878	10,441	5,920
Schools	543	413	235	543	413	235

Notes: The reduced form models regress the outcome variables (Teacher Experience, Tenure) on eligibility for the program since 2005. The IV-models use eligibility for the program since 2005 as instrument for participation in the program. All models control for grade, year and a quadratic of school size. Data used are at the grade by year level. The three discontinuity samples (+/- 1.5; +/- 1.0; +/- 0.5) are based on standard deviations of the poverty score across the cut-off. Standard errors adjusted for clustering at the school X year level. ***, **, * statistically significant at the 1, 5 or 10 %-level.

Table 3. Estimates of the effect of the CSCC program on student outcomes 2009-2013 (pooled data)

	Insufficient Attendance			Grade Retention			Drop Out		
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
Discontinuity sample	+/- 1.5	+/- 1.0	+/- 0.5	+/- 1.5	+/- 1.0	+/- 0.5	+/- 1.5	+/- 1.0	+/- 0.5
Reduced form									
1st-order polynomial	-0.052	-0.153***	-0.093	0.014	-0.007	0.132*	0.039	-0.056	-0.066
	(0.046)	(0.055)	(0.076)	(0.038)	(0.047)	(0.067)	(0.061)	(0.066)	(0.045)
2nd-order polynomial	-0.252***	-0.108	0.080	-0.081	0.001	-0.008	-0.079	-0.015	-0.155
	(0.071)	(0.086)	(0.115)	(0.060)	(0.075)	(0.107)	(0.061)	(0.055)	(0.140)
IV-estimates									
1st-order polynomial	-0.095	-0.316***	-0.255	0.026	-0.011	0.361*	0.068	-0.115	-0.183
	(0.084)	(0.117)	(0.214)	(0.069)	(0.097)	(0.187)	(0.110)	(0.138)	(0.128)
2nd-order polynomial	-0.620***	-0.346	0.278	-0.198	0.005	0.004	-0.189	-0.056	-0.619
	(0.194)	(0.268)	(0.464)	(0.152)	(0.224)	(0.427)	(0.157)	(0.168)	(0.608)
Observations	15288	11577	6599	15297	11583	6602	15289	11576	6599
Schools	543	413	235	543	413	235	543	413	235

Notes: The reduced form models regress the outcome variables (Insufficient Attendance, Grade Retention, Drop Out) on eligibility for the program since 2005. The IV-models use eligibility for the program since 2005 as instrument for participation in the program. All models control for grade, year and a quadratic of school size. Data used are at the grade by year level. The three discontinuity samples (+/- 1.5; +/- 1.0; +/- 0.5) are based on standard deviations of the poverty score across the cut-off. Standard errors adjusted for clustering at the school X year level. ***, **, * statistically significant at the 1, 5 or 10 %-level.

Table 4. Estimates of the effect of the CSCC program on Math and Language Test Scores

	(1)	(2)	(3)	(4)	(5)	(6)
Discontinuity sample	+/- 1.5	+/- 1.0	+/- 0.5	+/- 1.5	+/- 1.0	+/- 0.5
Math		**2009**			**Difference-in-differences**	
1st-order polynomial	-0.084	-0.228	0.109	-0.069	-0.082	-0.190
	(0.147)	(0.172)	(0.252)	(0.091)	(0.114)	(0.146)
2nd-order polynomial	-0.126	-0.093	-0.208	-0.090	-0.071	-0.214
	(0.206)	(0.248)	(0.332)	(0.088)	(0.113)	(0.142)
Observations	2,641	1,474	880	9,877	7,239	4,110
Schools	95	63	37	267	202	116
Language		**2009**			**Difference-in-differences**	
1st-order polynomial	-0.094	-0.147	0.145	-0.081	-0.090	-0.131
	(0.120)	(0.152)	(0.193)	(0.078)	(0.098)	(0.125)
2nd-order polynomial	-0.035	-0.017	0.012	-0.098	-0.084	-0.145
	(0.177)	(0.195)	(0.222)	(0.077)	(0.097)	(0.123)
Observations	2,641	1,474	880	10,116	7,381	4,187
Schools	95	63	37	267	202	116

Notes: Reduced form models have been estimated for 2009. These models regress the outcome variable on program eligibility since 2005. All models control for age and gender. The difference-in-difference estimates compare test scores before and after 2005 and across the cutoff for program eligibility. The number of schools in the difference-in-differences analysis is smaller than the sum of schools over the years because some schools participated in more than one year. The discontinuity samples (+/- 1.5; +/- 1.0; +/- 0.5) are based on standard deviations of the poverty score across the cut-off. Standard errors adjusted for clustering at the school level. ***, **, * statistically significant at the 1, 5 or 10 %-level.

Table 5. Reduced form estimates of the effect of the CSCC program on 'relevant experience'

	(1)	(2)	(3)	(4)	(5)	(6)
Discontinuity sample	+/- 1.5	+/- 1.0	+/- 0.5	+/- 1.5	+/- 1.0	+/- 0.5
Teacher experience:		**2006**			**2009-2013**	
0-4 years	-0.010	0.023	-0.019	-0.058***	-0.083***	-0.076**
	(0.045)	(0.055)	(0.071)	(0.020)	(0.023)	(0.031)
5-9 years	-0.031	-0.034	-0.045	-0.035**	-0.055***	-0.095***
	(0.039)	(0.049)	(0.070)	(0.016)	(0.020)	(0.027)
10-14 years	0.086***	0.097***	0.078	0.057***	0.060***	0.001
	(0.032)	(0.037)	(0.050)	(0.015)	(0.019)	(0.026)
15-19 years	0.025	-0.008	0.014	0.038***	0.064***	0.096***
	(0.027)	(0.032)	(0.043)	(0.013)	(0.015)	(0.021)
20-24 years	-0.017	0.005	0.040	-0.001	-0.008	0.015
	(0.022)	(0.027)	(0.036)	(0.010)	(0.013)	(0.019)
25-29 years	-0.057**	-0.063**	-0.046	0.007	0.008	0.016
	(0.023)	(0.027)	(0.037)	(0.008)	(0.010)	(0.013)
30-34 years	-0.012	-0.027	0.009	-0.010	-0.007	0.002
	(0.021)	(0.025)	(0.036)	(0.008)	(0.010)	(0.014)
35-40 years	-0.002	-0.003	-0.013	0.006	0.006	0.012*
	(0.005)	(0.006)	(0.008)	(0.004)	(0.005)	(0.007)
Observations	5,974	4,396	2,459	29,099	21,146	13,878

Notes: Based on data from individual teachers. Each estimate is based on a linear probability model that uses the teacher experience category, for instance '0-4 years', as dependent variable. This dependent variable is regressed on program eligibility, a quadratic specification of the forcing variable, grade, year, school size and school size squared. The three discontinuity samples (+/- 1.5; +/- 1.0; +/- 0.5) are based on standard deviations of the poverty score across the cut-off. Standard errors adjusted for clustering at the school X year level.

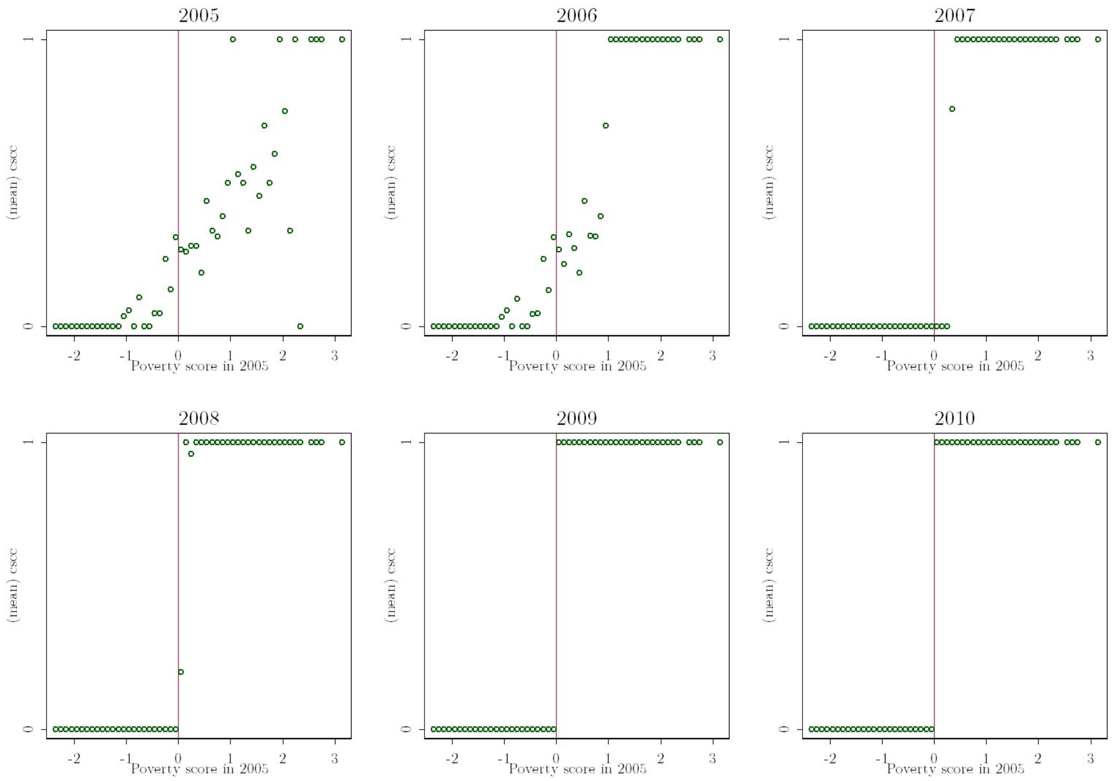
Table 6. Reduced form estimates of the effect of the CSCC program on student outcomes 2009-2013 for schools with little or many rookie teachers (less than 5 years of experience) in 2005

	Insufficient Attendance			**Grade Retention**			**Drop Out**		
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
Discontinuity sample	+/- 1.5	+/- 1.0	+/- 0.5	+/- 1.5	+/- 1.0	+/- 0.5	+/- 1.5	+/- 1.0	+/- 0.5
All grades									
Schools with above median % of rookie teachers	-0.200***	-0.300***	-0.252**	-0.180***	-0.178***	0.031	-0.085**	-0.170***	-0.118**
	(0.069)	(0.082)	(0.105)	(0.054)	(0.066)	(0.089)	(0.042)	(0.047)	(0.056)
Observations	5,931	4,665	2,981	5,934	4,668	2,982	5,928	4,662	2,980
Schools with below median % of rookie teachers	0.047	-0.084	0.096	0.159***	0.140**	0.220**	0.032	-0.012	-0.010
	(0.069)	(0.084)	(0.118)	(0.057)	(0.071)	(0.108)	(0.042)	(0.046)	(0.065)
Observations	7,931	5,750	2,979	7,937	5,753	2,981	7,935	5,752	2,980
Grades 1-2									
Schools with above median % of rookie teachers	-0.340***	-0.439***	-0.451***	-0.404***	-0.398***	-0.041	-0.088	-0.220**	-0.167*
	(0.098)	(0.112)	(0.138)	(0.088)	(0.105)	(0.140)	(0.078)	(0.087)	(0.100)
Observations	1,978	1,556	994	1,978	1,556	994	1,978	1,556	994
Schools with below median % of rookie teachers	0.038	-0.141	0.142	0.141	0.167	0.249	0.101	-0.013	0.105
	(0.093)	(0.113)	(0.160)	(0.093)	(0.115)	(0.172)	(0.062)	(0.076)	(0.101)
Observations	2,645	1,917	993	2,645	1,917	993	2,643	1,916	992

Notes: Student outcomes are regressed on eligibility for the program since 2005. The first rows only include schools that had many rookie teachers in 2005 (schools with above median proportion of rookie teachers in 2005). The next rows only include schools that had little rookie teachers in 2005. All models control for grade, year and a quadratic of school size, and use a first order polynomial of poverty score. Data used are at the grade by year level. The three discontinuity samples (+/- 1.5; +/- 1.0; +/- 0.5) are based on standard deviations of the poverty score across the cut-off. Standard errors adjusted for clustering at the school X year level. ***, **, * statistically significant at the 1, 5 or 10 %-level.

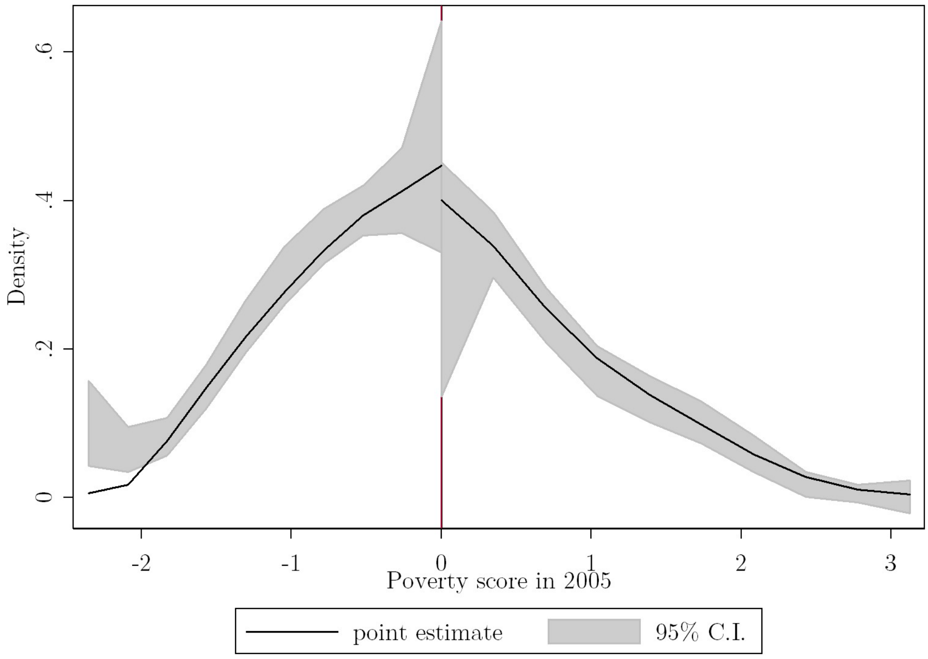
Online **Appendix Figures and Tables**

Figure A.1 Participation of schools in the CSCC-program by poverty score 2005-2010



Notes: These figures show program participation between 2005 and 2010 by score on the poverty index. Schools with poverty scores above zero are eligible for the program. Each dot represents the mean of the dependent variable (program participation) for schools located within a bin of width 0.1 of the poverty score.

Figure A.2 Density of the forcing variable across the cut-off for program eligibility



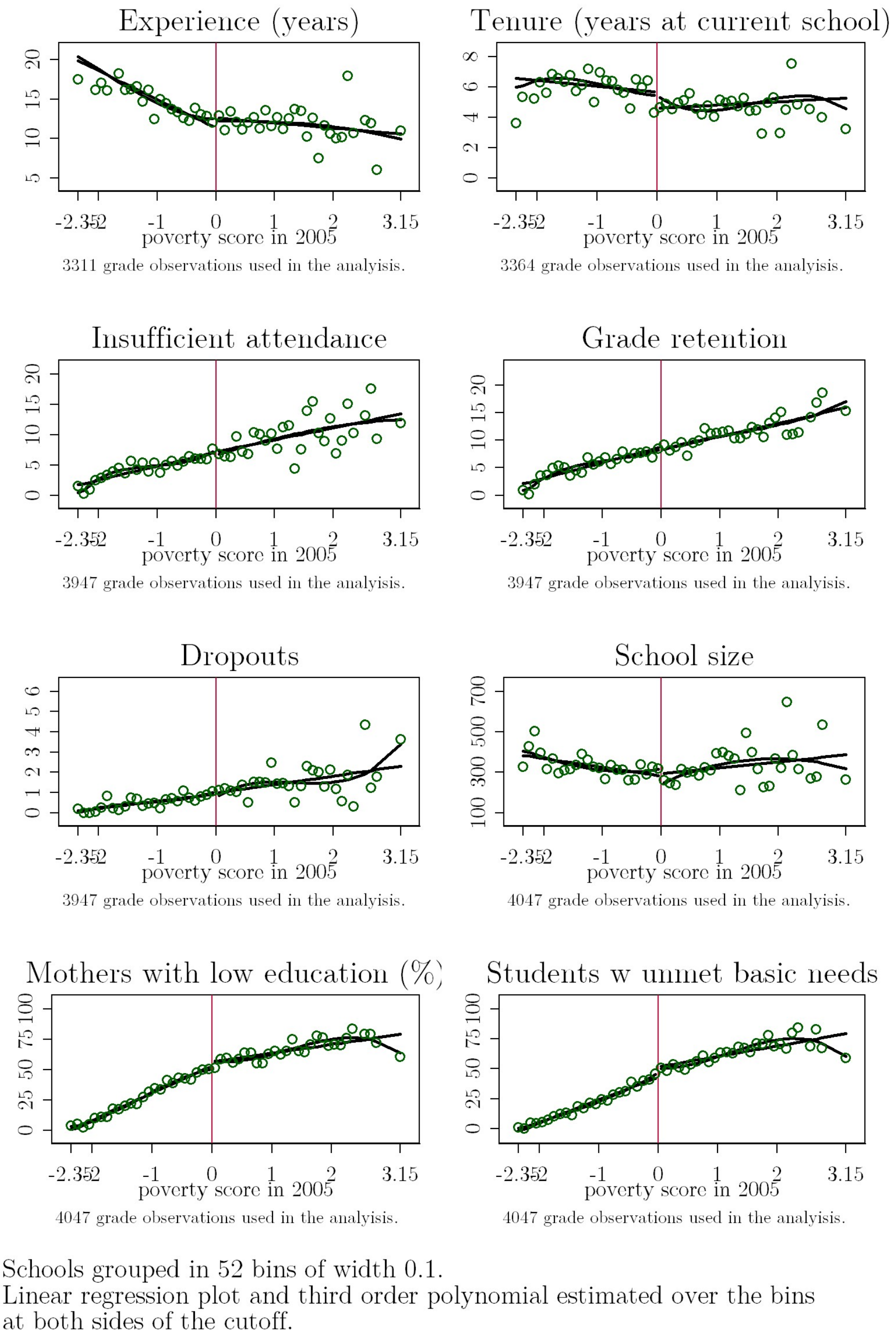
Manipulation tests:

1. Conventional T=-0.61 P-value (0.54)

2. Robust T=-1.64 P-value (0.10)

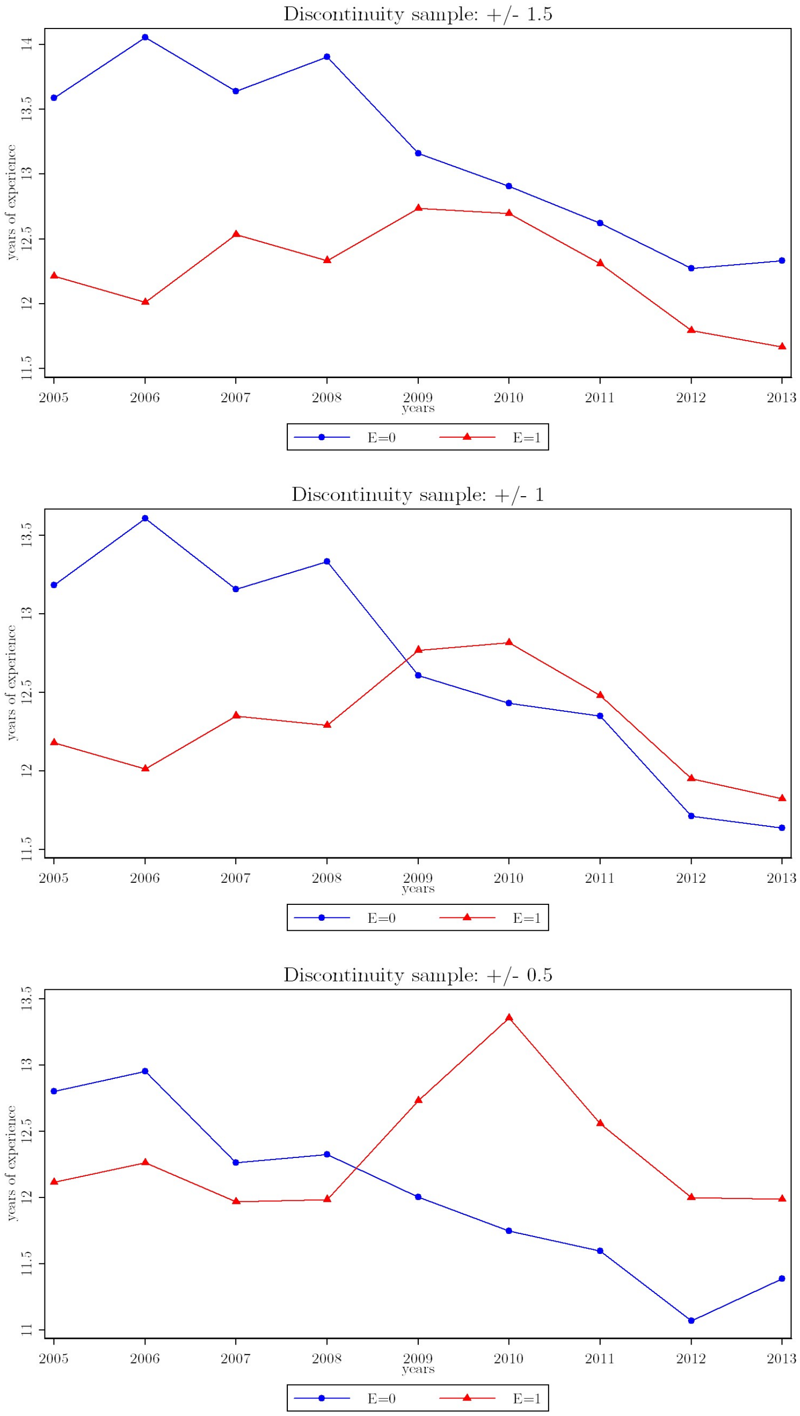
Note: The figure shows the density of the forcing variable across the cutoff. The test for the maninpulation across the cut-off use the methods by Cattaneo, Jansson and Ma (2017).

Figure A.3 Balancing tests for pre-treatment outcomes and characteristics in 2005



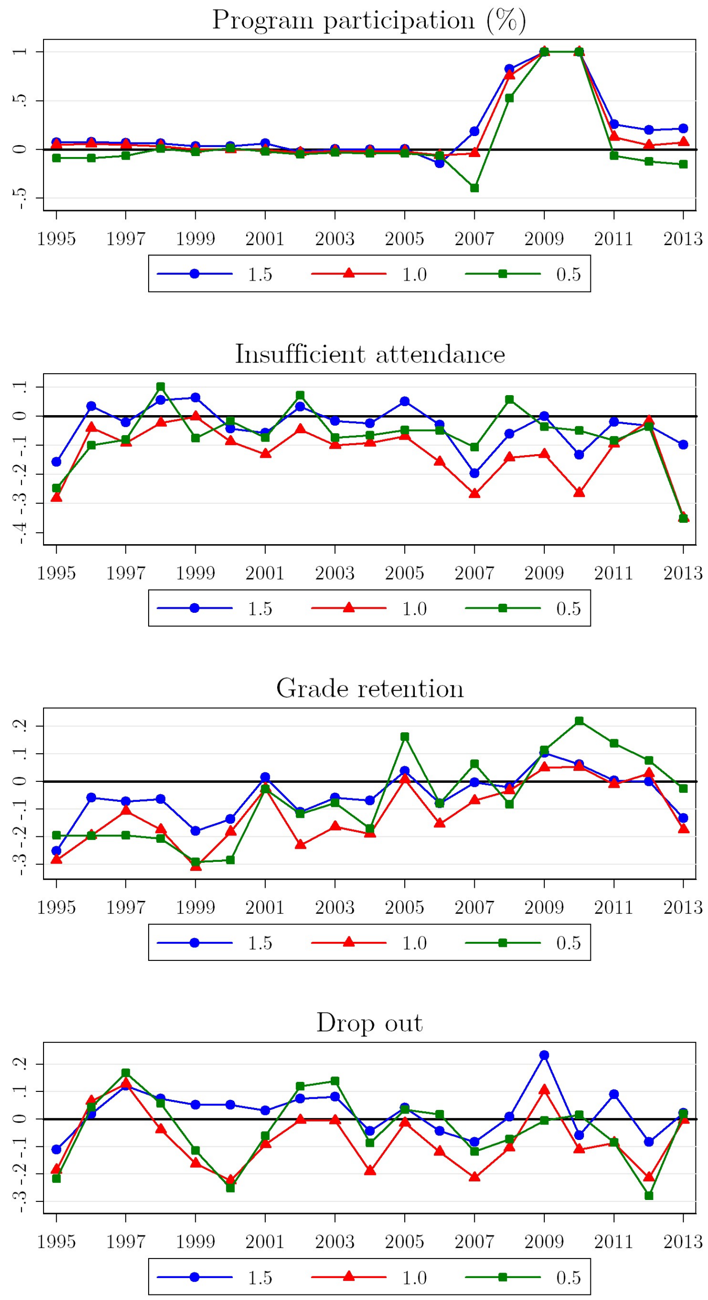
Notes: Each dot in the figure represents the mean of of pre-treatment outcomes or covariates for schools located within a bin of width 0.1 of the poverty score. The figures use a linear and cubic fit for the regression lines at both sides of the cut-off for eligibility for the program.

Figure A.4 Teacher experience 2005-2013 for three discontinuity samples



Notes: The figures show the mean of teacher experience for schools that are eligible for the CSCC-program and schools that are not eligible for the period 2005-2013 and for three discontinuity sample around the cut-off.

Figure A.5 Differences in program participation, insufficient attendance, grade retention and drop out for schools at the cut-off using three discontinuity samples 1995-2013



Notes: The figures show point estimates from regression-discontinuity models with a linear specification of the forcing variable and controlling for grade, year and a quadratic of school size. The three discontinuity samples used are +/- 1.5, +/- 1.0 and +/- 0.5 points of the running variable.

Table A.1 Teachers' salaries (base plus additional salary)

	Payment Categories		
Year	#1	#4	#7
1997	9,365	11,235	16,827
1998	9,453	11,346	17,004
1999	10,787	12,144	18,242
2000	10,697	12,044	18,090
2001	10,494	11,814	17,746
2002	11,101	12,519	18,300
2003	8,990	10,138	14,819
2004	8,983	10,186	14,784
2005	9,230	10,424	14,512
2006	9,728	11,033	15,389
2007	10,441	11,927	16,746
2008	11,760	13,568	19,221
2009	12,777	14,918	21,201
2010	13,001	15,182	21,593

Notes: Teacher salaries by year and payment category. The payment scale includes seven categories. A new teacher starts in category #1 and moves to category #2 after four years. Hence, after 12 years of work, she can reach the 4th category: this is equal to an increase of 15% in her base salary. The columns show nominal wage (with food complements) for 20hs teachers in Levels 1, 4 and 7 of the payment scale, in constant Uruguayan pesos of February 2011. Data from January in each year. Source: Area de Estadística y Análisis- Dirección Sectorial de Programación y Presupuesto - CODICEN- ANEP

Table A.2 Estimates of the effect of being eligible for the CSCC-program on the use of non-salary components

	(1)	(2)	(3)	(4)	(5)	(6)
Non-salary component	% kids having lunch at school	has library room	computers for educational use	Number of computers	study books per capita - 1st grade	School has community teacher
Program eligibility	0.006	-0.066	0.145	0.612	-0.107	-0.016
	(0.065)	(0.132)	(0.125)	(0.830)	(0.315)	(0.127)
Poverty-index	0.425***	-0.173	-0.078	-1.095	-0.216	0.611**
	(0.157)	(0.314)	(0.306)	(1.733)	(0.815)	(0.299)
CSCC*Poverty-index	-0.565**	0.150	-0.224	-0.440	1.171	0.099
	(0.228)	(0.461)	(0.451)	(2.815)	(1.282)	(0.418)
Constant	0.624***	0.329***	0.690***	1.954***	1.227***	0.642***
	(0.040)	(0.085)	(0.083)	(0.451)	(0.234)	(0.081)
Observations	237	220	214	214	197	237
R-squared	0.059	0.016	0.007	0.005	0.005	0.141

Notes: Each column shows estimates of regression discontinuity models that regress a non-salary component on program eligibility and a linear specification of poverty score for both sides of the cut-off. The regression discontinuity models is specified as in Equation (1) in Section 4 using a discontinuity sample of 0.5 (schools with absolute poverty scores smaller than 0.5). Dates of measurement: students having lunch (2010), library (2008), computers (2009), study books (2011), community teachers (2010). Robust standard errors in parentheses. *** p<0.01, ** p<0.05, * p<0.1.

Table A.3 Means of outcome variables 2005-2013 (discontinuity sample +/- 0.50)

Years		2005	2006	2007	2008	2009	2010	2011	2012	2013
Panel A: Teachers										
Experience (years)	Non elegible	12.8	13.0	12.3	12.3	12.0	11.7	11.6	11.1	11.4
	Elegible	12.1	12.3	12.0	12.0	12.7	13.4	12.6	12.0	12.0
Tenure (years)	Non elegible	5.6	5.3	5.3	5.3	5.0	4.6	4.5	4.0	4.5
	Elegible	4.8	5.0	4.7	4.9	5.0	4.9	4.8	4.4	4.8
Panel B: Students										
Insufficient attendance (%)	Non elegible	6.4	5.4	7.7	6.6	9.8	7.3	5.5	6.2	8.1
	Elegible	7.5	5.8	7.3	6.6	10.9	6.9	6.2	7.3	7.4
Grade retention (%)	Non elegible	7.7	8.1	7.7	6.3	5.9	5.8	5.7	5.4	5.7
	Elegible	8.7	8.1	8.1	6.7	6.9	6.2	5.7	5.9	4.9
Dropout (%)	Non elegible	0.8	0.9	1.1	0.9	1.2	1.0	0.8	1.1	0.6
	Elegible	1.1	0.9	1.1	0.9	2.7	1.0	1.1	1.0	0.8

Notes: Means of outcomes for teachers and students for 2005-2013

Table A.4 Estimates of the pre-treatment balance in 2005

Discontinuity sample	+/- 1.5	+/- 1.0	+/- 0.5
	(1)	(2)	(3)
Teacher Experience (in years)			
CSCC Program	-0,033	0,007	0,028
	(0.82)	(0.963)	(1.459)
Observations	2.796	2.113	1.211
Tenure (in years at current school)			
CSCC Program	-0,67	-0,34	-0,892
	(0.43)	(0.507)	(0.73)
Observations	2.856	2.168	1.234
Insufficient Attendance			
CSCC Program	0,247	-0,833	-0,898
	(0.704)	(0.832)	(1.171)
Observations	3.371	2.573	1.470
Grade Retention			
CSCC Program	0,074	-0,228	1,079
	(0.674)	(0.794)	(1.129)
Observations	3.371	2.573	1.470
Dropout			
CSCC Program	0,094	-0,141	-0,003
	(0.205)	(0.26)	(0.362)
Observations	3.371	2.573	1.470
School size (number of students in primary education)			
CSCC Program	-60.428**	-74.557**	-100.195**
	(27.817)	(32.038)	(43.929)
Observations	562	429	245
Mothers with primary education or less (%)			
CSCC Program	2,034	3.891*	1,311
	(1.796)	(2.075)	(2.845)
Observations	562	429	245
Children with unmet basic needs (%)			
CSCC Program	1,927	1,713	3,768
	(1.94)	(2.411)	(3.448)
Observations	562	429	245

Notes: Coefficients are obtained from regressions of the outcome or covariate in 2005 on the poverty score, an indicator for program eligibility, their interaction and grade dummies. CSCC-program is a dummy for program eligibility. Standard errors clustered at the school level. *** p<0.01, ** p<0.05, * p<0.1

Table A.5 Sensitivity analysis: reduced form estimates of the effect of the CSCC program on experience and tenure 2009-2013 (pooled data)

	Teacher experience			Tenure		
	(1)	(2)	(3)	(4)	(5)	(6)
Discontinuity sample	+/- 1.5	+/- 1.0	+/- 0.5	+/- 1.5	+/- 1.0	+/- 0.5
No controls						
1st-order	1.930***	2.011***	2.503***	0.371**	0.743***	0.611**
	(0.312)	(0.368)	(0.516)	(0.188)	(0.217)	(0.289)
2nd-order	2.256***	2.760***	4.181***	0.915***	0.560*	1.494***
	(0.463)	(0.567)	(0.773)	(0.258)	(0.309)	(0.388)
Cubic	2.571***	2.846***	4.551***	0.915***	1.237***	2.546***
	(0.603)	(0.718)	(0.963)	(0.333)	(0.389)	(0.450)
Local Polynomial						
1st-order	2.949***	3.711***	3.698***	1.592***	2.080***	2.159***
	(0.955)	(1.228)	(1.297)	(0.481)	(0.478)	(0.493)
2nd-order	3.959***	3.847***	3.522*	2.091***	2.227***	2.240***
	(1.370)	(1.483)	(1.842)	(0.577)	(0.581)	(0.614)
Additonal controls						
1st-order	1.658***	1.570***	1.706***	0.414**	0.761***	0.697**
	(0.302)	(0.356)	(0.505)	(0.190)	(0.219)	(0.289)
2nd-order	1.709***	2.134***	3.144***	0.997***	0.647**	1.532***
	(0.448)	(0.549)	(0.732)	(0.259)	(0.309)	(0.374)
Observations	13,749	10,341	5,868	13,878	10,441	5,920
Controlling for pre-treatment dependent						
1st-order	1.822***	1.709***	1.930***	0.551***	0.892***	0.907***
	(0.283)	(0.329)	(0.453)	(0.183)	(0.209)	(0.278)
2nd-order	1.792***	2.195***	3.497***	1.149***	0.784***	1.384***
	(0.408)	(0.494)	(0.661)	(0.249)	(0.298)	(0.373)
Observations	12,902	9,645	5,441	13,028	9,741	5,492
School level data						
1st-order	1.571***	1.356***	1.701***	0.329	0.671***	0.462
	(0.337)	(0.407)	(0.559)	(0.202)	(0.247)	(0.350)
2nd-order	1.523***	1.956***	2.237**	0.937***	0.544	1.079**
	(0.510)	(0.623)	(0.869)	(0.305)	(0.378)	(0.542)
Observations	2,482	1,881	1,075	2,495	1,892	1,08

Notes: Data used are at the grade level. Standard errors adjusted for clustering at the school X year level. Local Polynomial RD point estimators as developed in Calonico et al. (2014). Additional controls are 'mothers with primary education or less in 2005' and 'students with unmet basic needs in 2005'. Pre-treatment dependent for respectively experience or tenure are taken as most recent available from 2003-2005. ***, **, * statistically significant at the 1, 5 or 10 %-level.

Table A.6 Reduced form estimates of the effect of being eligible for the CSCC-program on teacher experience and tenure by year 2006-2013

	Teacher experience			Tenure		
Discontinuity sample	+/- 1.5	+/- 1.0	+/- 0.5	+/- 1.5	+/- 1.0	+/- 0.5
	(1)	(2)	(3)	(4)	(5)	(6)
year = 2006	-0.287	-0.038	-0.088	-0.153	0.450	0.213
	(0.810)	(0.961)	(1.355)	(0.501)	(0.585)	(0.840)
Observations	2,922	2,220	1,242	2,961	2,251	1,263
year = 2007	-0.475	-0.613	-0.478	-0.434	-0.198	-0.861
	(0.770)	(0.900)	(1.307)	(0.478)	(0.541)	(0.788)
Observations	2,933	2,249	1,270	2,955	2,262	1,283
year = 2008	-0.022	-0.492	-0.527	-0.309	0.025	-0.715
	(0.792)	(0.952)	(1.479)	(0.477)	(0.547)	(0.809)
Observations	2,971	2,235	1,261	2,996	2,254	1,279
year = 2009	1.228*	1.185	2.462**	-0.189	0.438	-0.118
	(0.738)	(0.878)	(1.243)	(0.445)	(0.501)	(0.663)
Observations	2,814	2,114	1,196	2,855	2,150	1,211
year = 2010	2.279***	2.640***	3.099***	0.340	0.832*	0.790
	(0.670)	(0.782)	(1.163)	(0.443)	(0.500)	(0.708)
Observations	2,949	2,224	1,262	2,974	2,245	1,275
year = 2011	1.712***	1.658**	2.235**	0.586	0.930*	0.816
	(0.634)	(0.745)	(0.986)	(0.405)	(0.479)	(0.640)
Observations	2,780	2,107	1,207	2,804	2,117	1,208
year = 2012	1.637**	1.248	0.815	0.778*	0.679	0.740
	(0.669)	(0.796)	(1.079)	(0.414)	(0.479)	(0.617)
Observations	2,732	2,067	1,155	2,749	2,082	1,160
year = 2013	1.483**	0.684	0.405	0.532	0.825*	1.225*
	(0.672)	(0.790)	(1.082)	(0.423)	(0.496)	(0.667)
Observations	2,474	1,829	1,048	2,496	1,847	1,066

Notes: Estimates from models using a linear specification of the forcing variable and controls for grade and a quadratic of school size. Data used are at the grade level. Standard errors adjusted for clustering at the school level. ***, **, * statistically significant at the 1, 5 or 10 %-level.

Table A.7 Reduced form estimates of the effect of being eligible for the CSCC-program on insufficient attendance of students by year 2006-2013

Discontinuity sample	+/- 1.5	+/- 1.0	+/- 0.5
	(1)	(2)	(3)
year = 2006	-0.028	-0.152	-0.048
	(0.072)	(0.093)	(0.123)
# Observations	3,365	2,567	1,470
year = 2007	-0.190**	-0.259**	-0.102
	(0.091)	(0.106)	(0.153)
# Observations	3,358	2,560	1,464
year = 2008	-0.059	-0.137	0.055
	(0.091)	(0.108)	(0.148)
# Observations	3,309	2,517	1,434
year = 2009	-0.000	-0.127	-0.035
	(0.124)	(0.148)	(0.215)
# Observations	3,252	2,475	1,409
year = 2010	-0.127	-0.255**	-0.048
	(0.097)	(0.119)	(0.169)
# Observations	3,238	2,464	1,403
year = 2011	-0.019	-0.092	-0.081
	(0.082)	(0.093)	(0.128)
# Observations	3,081	2,337	1,331
year = 2012	-0.032	-0.015	-0.035
	(0.095)	(0.111)	(0.145)
# Observations	2,985	2,265	1,278
year = 2013	-0.094	-0.337**	-0.339
	(0.148)	(0.166)	(0.228)
# Observations	2,732	2,036	1,178

Notes: Estimates from models using a linear specification of the forcing variable and controls for grade and a quadratic of school size. Data used are at the grade level. Standard errors adjusted for clustering at the school level. ***, **, * statistically significant at the 1, 5 or 10 %-level.

Table A.8 Reduced form estimates of the effect of being eligible for the CSCC program on grade retention by year 2006-2013

Discontinuity sample	+/- 1.5	+/- 1.0	+/- 0.5
	(5)	(6)	(7)
year = 2006	-0.092	-0.179*	-0.093
	(0.080)	(0.097)	(0.148)
# Observations	3,365	2,567	1,470
year = 2007	-0.004	-0.080	0.074
	(0.088)	(0.107)	(0.160)
# Observations	3,358	2,560	1,464
year = 2008	-0.025	-0.038	-0.097
	(0.090)	(0.111)	(0.180)
# Observations	3,309	2,517	1,434
year = 2009	0.121	0.058	0.134
	(0.081)	(0.098)	(0.142)
# Observations	3,257	2,477	1,410
year = 2010	0.073	0.061	0.256
	(0.089)	(0.107)	(0.165)
# Observations	3,238	2,464	1,403
year = 2011	0.003	-0.011	0.160
	(0.076)	(0.089)	(0.127)
# Observations	3,081	2,337	1,331
year = 2012	0.000	0.035	0.090
	(0.087)	(0.108)	(0.148)
# Observations	2,985	2,265	1,278
year = 2013	-0.154*	-0.202**	-0.030
	(0.081)	(0.098)	(0.135)
# Observations	2,736	2,040	1,180

Notes: Estimates from models using a linear specification of the forcing variable and controls for grade and a quadratic of school size. Data used are at the grade level. Standard errors adjusted for clustering at the school level. ***, **, * statistically significant at the 1, 5 or 10 %-level.

Table A.9 Reduced form estimates of the effect of being eligible for the CSCC program on the drop out by year 2006-2013

Discontinuity sample	+/- 1.5	+/- 1.0	+/- 0.5
	(9)	(10)	(11)
year = 2006	-0.043	-0.119*	0.018
	(0.052)	(0.067)	(0.082)
# Observations	3,364	2,566	1,470
year = 2007	-0.083	-0.214***	-0.118
	(0.064)	(0.079)	(0.092)
# Observations	3,356	2,558	1,463
year = 2008	0.008	-0.104	-0.072
	(0.062)	(0.068)	(0.082)
# Observations	3,309	2,517	1,434
year = 2009	0.233	0.105	-0.005
	(0.295)	(0.311)	(0.165)
# Observations	3,249	2,47	1,407
year = 2010	-0.059	-0.111	0.015
	(0.066)	(0.081)	(0.104)
# Observations	3,238	2,464	1,403
year = 2011	0.090	-0.086	-0.085
	(0.067)	(0.083)	(0.115)
# Observations	3,081	2,337	1,331
year = 2012	-0.083	-0.213***	-0.279***
	(0.071)	(0.082)	(0.090)
# Observations	2,985	2,265	1,278
year = 2013	0.023	-0.004	0.019
	(0.055)	(0.067)	(0.088)
# Observations	2,736	2,040	1,180

Notes: Estimates from models using a linear specification of the forcing variable and controls for grade and a quadratic of school size. Data used are at the grade level. Standard errors adjusted for clustering at the school level. ***, **, * statistically significant at the 1, 5 or 10 %-level.

Table A.10 Sensitivity analysis: reduced form estimates of the effect of the CSCC program on student outcomes 2009-2013

	Insufficient Attendance			Grade Retention			Drop Out		
Discontinuity sample	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
	+/- 1.5	+/- 1.0	+/- 0.5	+/- 1.5	+/- 1.0	+/- 0.5	+/- 1.5	+/- 1.0	+/- 0.5
Cubic	-0.040	0.014	0.370**	0.079	0.082	0.155	-0.093	-0.174*	-0.019
	(0.094)	(0.113)	(0.144)	(0.083)	(0.102)	(0.136)	(0.057)	(0.099)	(0.089)
Local Polynomial									
1st-order	0.191	0.221	0.269	0.071	0.057	0.027	-0.029	0.011	0.109
	(0.171)	(0.184)	(0.186)	(0.130)	(0.164)	(0.209)	(0.069)	(0.085)	(0.101)
2nd-order	0.287	0.314	0.631***	0.102	0.065	0.020	-0.087	0.048	0.092
	(0.186)	(0.199)	(0.195)	(0.183)	(0.210)	(0.240)	(0.105)	(0.103)	(0.108)
Additonal controls									
1st-order	-0.045	-0.138**	-0.109	0.019	0.002	0.125*	0.036	-0.061	-0.070
	(0.046)	(0.054)	(0.076)	(0.038)	(0.046)	(0.067)	(0.057)	(0.058)	(0.047)
2nd-order	-0.241***	-0.124	0.055	-0.075	-0.007	-0.017	-0.082	-0.018	-0.152
	(0.070)	(0.085)	(0.113)	(0.059)	(0.074)	(0.106)	(0.057)	(0.057)	(0.131)
Observations	15288	11577	6599	15297	11583	6602	15289	11576	6599
Pre-treatment dependent as control									
1st-order	-0.063	-0.140***	-0.079	0.008	-0.008	0.111*	0.035	-0.057	-0.067
	(0.043)	(0.052)	(0.073)	(0.037)	(0.045)	(0.065)	(0.062)	(0.066)	(0.045)
2nd-order	-0.210***	-0.089	0.126	-0.084	-0.018	0.003	-0.078	-0.022	-0.155
	(0.067)	(0.081)	(0.110)	(0.058)	(0.072)	(0.103)	(0.061)	(0.055)	(0.140)
Observations	15,284	11,573	6,599	15,293	11,579	6,602	15,285	11,572	6,599
School level data									
1st-order	-0.056	-0.156***	-0.095	0.014	-0.006	0.132**	0.052	-0.061	-0.067
	(0.048)	(0.057)	(0.076)	(0.038)	(0.046)	(0.062)	(0.060)	(0.080)	(0.108)
2nd-order	-0.253***	-0.107	0.082	-0.082	-0.000	-0.008	-0.102	-0.057	-0.153
	(0.073)	(0.087)	(0.119)	(0.057)	(0.070)	(0.097)	(0.091)	(0.122)	(0.168)
Observations	2,552	1,933	1,101	2,552	1,933	1,101	2,552	1,933	1,101

Notes: see Table A.5.

Table A.11 Estimates of the effect of the CSCC program on student outcomes by grade 2009-2013

Grades 1-2		**Insufficient Attendance**			**Grade Retention**			**Drop Out**	
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
Discontinuity sample	+/- 1.5	+/- 1.0	+/- 0.5	+/- 1.5	+/- 1.0	+/- 0.5	+/- 1.5	+/- 1.0	+/- 0.5
Reduced form									
1st-order polynomial	-0.111*	-0.235***	-0.167	-0.091	-0.110	0.091	0.071	-0.106	-0.038
	(0.065)	(0.075)	(0.103)	(0.064)	(0.076)	(0.105)	(0.073)	(0.083)	(0.072)
2nd-order polynomial	-0.390***	-0.273**	0.013	-0.212**	-0.058	-0.207	-0.134	0.027	-0.039
	(0.096)	(0.115)	(0.156)	(0.095)	(0.114)	(0.162)	(0.092)	(0.112)	(0.165)
IV-estimates									
1st-order polynomial	-0.201*	-0.487***	-0.458	-0.163	-0.227	0.254	0.124	-0.215	-0.104
	(0.117)	(0.162)	(0.292)	(0.115)	(0.159)	(0.288)	(0.130)	(0.172)	(0.200)
2nd-order polynomial	-0.960***	-0.847**	-0.009	-0.526**	-0.179	-0.774	-0.321	0.059	-0.172
	(0.268)	(0.387)	(0.625)	(0.245)	(0.344)	(0.745)	(0.233)	(0.338)	(0.657)
Observations	5,100	3,862	2,200	5,100	3,862	2,200	5,098	3,861	2,199
Grades 3-6									
Reduced form									
1st-order polynomial	-0.022	-0.112**	-0.055	0.067*	0.045	0.152**	0.024	-0.032	-0.081*
	(0.044)	(0.053)	(0.075)	(0.035)	(0.043)	(0.062)	(0.062)	(0.066)	(0.047)
2nd-order polynomial	-0.183***	-0.025	0.113	-0.016	0.031	0.091	-0.052	-0.036	-0.212
	(0.069)	(0.084)	(0.113)	(0.055)	(0.071)	(0.102)	(0.062)	(0.050)	(0.143)
IV-estimates									
1st-order polynomial	-0.042	-0.230**	-0.152	0.120*	0.098	0.415**	0.041	-0.064	-0.222*
	(0.080)	(0.112)	(0.207)	(0.062)	(0.090)	(0.175)	(0.112)	(0.138)	(0.134)
2nd-order polynomial	-0.449**	-0.095	0.421	-0.034	0.097	0.391	-0.123	-0.112	-0.841
	(0.181)	(0.253)	(0.476)	(0.138)	(0.211)	(0.417)	(0.156)	(0.153)	(0.652)
Observations	10,188	7,715	4,399	10,197	7,721	4,402	10,191	7,715	4,400

Notes: The models from Table 3 are now estimated separately for grade 1-2 and for grades 3-6. All specifications as in Table 3.

Table A.12 Estimates of the effect of the CSCC program on attendance 2009-2013

	All grades			Grades 1-2			Grades 3-6		
Discontinuity sample	+/- 1.5	+/- 1.0	+/- 0.5	+/- 1.5	+/- 1.0	+/- 0.5	+/- 1.5	+/- 1.0	+/- 0.5
Reduced form estimates									
1st-order polynomial	0.038	0.149**	0.111	0.063	0.210***	0.118	0.025	0.118**	0.107
	(0.053)	(0.059)	(0.072)	(0.063)	(0.072)	(0.092)	(0.053)	(0.059)	(0.071)
2nd-order polynomial	0.251***	0.123	-0.000	0.334***	0.175	0.081	0.210***	0.097	-0.041
	(0.071)	(0.085)	(0.122)	(0.090)	(0.107)	(0.149)	(0.070)	(0.084)	(0.121)
IV-estimates									
1st-order polynomial	0.070	0.306**	0.303	0.116	0.433***	0.322	0.046	0.242*	0.294
	(0.095)	(0.126)	(0.204)	(0.114)	(0.156)	(0.259)	(0.095)	(0.125)	(0.202)
2nd-order polynomial	0.617***	0.391	0.036	0.821***	0.551	0.356	0.514***	0.310	-0.124
	(0.195)	(0.268)	(0.478)	(0.249)	(0.342)	(0.617)	(0.187)	(0.259)	(0.475)
Observations	15,297	11,583	6,602	5,100	3,862	2,200	10,197	7,721	4,402

Notes: The dependent variables ‘attendance’ is constructed from ‘insufficient attendance’, ‘dropout’ and additional information about the number of days that students attended school. Estimates are from models using a linear specification of the forcing variable. All models controls for grade and year. Data used are at the grade by year level. Standard errors adjusted for clustering at the school level. ***, **, * statistically significant at the 1, 5 or 10 %-level.

Table A.13 Estimates of the effect of the CSCC program on Math and Language Test Scores in 1999 and 2002

	(1)	(2)	(3)	(4)	(5)	(6)
Discontinuity sample	+/- 1.5	+/- 1.0	+/- 0.5	+/- 1.5	+/- 1.0	+/- 0.5
Math		**1999**			**2002**	
1st-order polynomial	0.109	0.158	0.375*	0.223	0.251	0.572
	(0.126)	(0.138)	(0.195)	(0.190)	(0.232)	(0.354)
2nd-order polynomial	0.220	0.233	0.316	0.378	0.527	0.459
	(0.165)	(0.190)	(0.313)	(0.301)	(0.371)	(0.439)
Observations	3,681	3,074	1,739	3,555	2,691	1,491
Schools	111	91	52	121	92	53
Language		**1999**			**2002**	
1st-order polynomial	-0.003	0.007	0.177	0.190	0.193	0.417
	(0.126)	(0.146)	(0.224)	(0.171)	(0.202)	(0.296)
2nd-order polynomial	0.036	-0.061	-0.172	0.284	0.408	0.545
	(0.181)	(0.229)	(0.332)	(0.254)	(0.308)	(0.404)
Observations	3,691	3,077	1,750	3,552	2,687	1,488
Schools	111	91	52	121	92	53

Notes: Reduced form models have been estimated for 1999 and 2002. These models regress the outcome variable on program eligibility since 2005. All models control for age and gender. The discontinuity samples (+/- 1.5; +/- 1.0; +/- 0.5) are based on standard deviations of the poverty score across the cut-off. Standard errors adjusted for clustering at the school level. ***, **, * statistically significant at the 1, 5 or 10 %-level.

Table A.14 Reduced form estimates of the effect of the CSCC program on student outcomes 2009-2013 for schools with many or little teachers with less than 4 years of tenure in 2005 (above or below median proportion of teachers with little tenure)

	Insufficient Attendance			**Grade Retention**			**Drop Out**		
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
Discontinuity sample	+/- 1.5	+/- 1.0	+/- 0.5	+/- 1.5	+/- 1.0	+/- 0.5	+/- 1.5	+/- 1.0	+/- 0.5
All grades									
Schools with above median %	-0.282***	-0.390***	-0.248**	-0.124**	-0.156**	-0.006	-0.087*	-0.176***	-0.126*
	(0.071)	(0.083)	(0.106)	(0.054)	(0.062)	(0.084)	(0.048)	(0.049)	(0.064)
Observations	6,241	5,12	3,232	6,246	5,124	3,234	6,239	5,117	3,231
Schools with below median %	0.152**	0.048	0.128	0.152**	0.145*	0.296***	0.038	-0.018	0.001
	(0.072)	(0.085)	(0.127)	(0.060)	(0.075)	(0.113)	(0.039)	(0.045)	(0.057)
Observations	7,621	5,295	2,728	7,625	5,297	2,729	7,624	5,297	2,729
Grades 1-2									
Schools with above median %	-0.388***	-0.525***	-0.412***	-0.266***	-0.300***	-0.039	-0.026	-0.209**	-0.070
	(0.098)	(0.111)	(0.141)	(0.087)	(0.099)	(0.133)	(0.074)	(0.087)	(0.107)
Observations	2,082	1,708	1,078	2,082	1,708	1,078	2,081	1,707	1,077
Schools with below median %	0.125	-0.002	0.148	0.087	0.123	0.308*	0.039	-0.061	0.036
	(0.098)	(0.116)	(0.167)	(0.095)	(0.117)	(0.169)	(0.063)	(0.080)	(0.093)
Observations	2,541	1,765	909	2,541	1,765	909	2,54	1,765	909

Notes: Student outcomes are regressed on eligibility for the program since 2005. The first rows only include schools that had many teachers with less than four years of tenure in 2005. The next rows only include schools that had little teachers with less than four years of tenure in 2005. All models control for grade, year and a quadratic of school size, and use a first order polynomial of poverty score. Data used are at the grade by year level. The three discontinuity samples (+/- 1.5; +/- 1.0; +/- 0.5) are based on standard deviations of the poverty score across the cut-off. Standard errors adjusted for clustering at the school X year level. ***, **, * statistically significant at the 1, 5 or 10 %-level.

Table A.15 Estimates of the effect of the CSCC program on tenure and experience at other schools

		Tenure			**Experience at other schools**	
	(1)	(2)	(3)	(4)	(5)	(6)
Discontinuity sample	+/- 1.5	+/- 1.0	+/- 0.5	+/- 1.5	+/- 1.0	+/- 0.5
Reduced form						
1st-order polynomial	0.403**	0.737***	0.657**	1.254***	0.819***	1.188***
	(0.191)	(0.220)	(0.297)	(0.248)	(0.298)	(0.435)
2nd-order polynomial	0.989***	0.640**	1.477***	0.728*	1.573***	1.726***
	(0.261)	(0.314)	(0.387)	(0.381)	(0.472)	(0.662)
Observations	13,878	10,441	5,920	13,749	10,341	5,868
Schools	543	413	235	543	413	235

Notes: Experience has been decomposed in ‘tenure’ and ‘experience at other schools’. Each experience component has been regressed on program eligibility since 2005 using the same reduced form model as in Table 2. The estimates in columns (1) to (3) are similar to those in columns (4) to (6) of Table 2. Data used from 2009-2013. Standard errors adjusted for clustering at the school level. ***, **, * statistically significant at the 1, 5 or 10 %-level.

Table A.16. Estimates of the effect of experience components on student outcomes 2009-2013 (pooled data)

	Insufficient Attendance			Grade Retention			Drop Out		
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
1st-order polynomial	+/- 1.5	+/- 1.0	+/- 0.5	+/- 1.5	+/- 1.0	+/- 0.5	+/- 1.5	+/- 1.0	+/- 0.5
Total experience	-0.003***	-0.003**	-0.001	-0.001	-0.001	0.003	-0.000	-0.000	-0.001
	(0.001)	(0.001)	(0.002)	(0.001)	(0.001)	(0.002)	(0.001)	(0.002)	(0.001)
Tenure	-0.011***	-0.007***	-0.004*	-0.002	-0.001	0.002	-0.003*	-0.001	-0.003**
	(0.002)	(0.002)	(0.002)	(0.001)	(0.002)	(0.002)	(0.002)	(0.002)	(0.002)
Experience at other schools	0.003*	0.000	0.002	-0.001	-0.001	0.003	0.002	-0.000	0.001
	(0.002)	(0.002)	(0.002)	(0.001)	(0.002)	(0.002)	(0.002)	(0.002)	(0.002)
2nd-order polynomial									
Total experience	-0.003***	-0.003**	-0.001	-0.001	-0.001	0.002	-0.000	-0.000	-0.001
	(0.001)	(0.001)	(0.002)	(0.001)	(0.001)	(0.002)	(0.001)	(0.002)	(0.001)
Tenure	-0.011***	-0.007***	-0.004*	-0.002*	-0.001	0.002	-0.003*	-0.001	-0.003**
	(0.002)	(0.002)	(0.002)	(0.001)	(0.002)	(0.002)	(0.002)	(0.002)	(0.002)
Experience at other schools	0.003*	0.001	0.002	-0.001	-0.001	0.002	0.001	0.000	0.001
	(0.002)	(0.002)	(0.002)	(0.001)	(0.002)	(0.002)	(0.002)	(0.002)	(0.002)
Fixed Effect models									
Total experience	-0.005***	-0.005***	-0.006***	-0.001	0.000	0.003*	-0.002	-0.002	-0.002
	(0.001)	(0.001)	(0.002)	(0.001)	(0.001)	(0.002)	(0.001)	(0.001)	(0.002)
Tenure	-0.011***	-0.010***	-0.013***	0.000	0.001	0.001	-0.002	-0.003	-0.005*
	(0.002)	(0.002)	(0.003)	(0.002)	(0.002)	(0.003)	(0.002)	(0.002)	(0.003)
Experience at other schools	-0.001	-0.002	-0.001	-0.001	-0.001	0.004**	-0.002	-0.002	-0.001
	(0.002)	(0.002)	(0.002)	(0.001)	(0.002)	(0.002)	(0.002)	(0.002)	(0.002)
Observations	13,295	9,982	5,659	13,304	9,988	5,662	13,296	9,981	5,659

Notes: Student outcomes are regressed on total experience or on the two components (tenure, experience at other schools). All models controls for grade, year, a quadratic of school size and the forcing variable which is allowed to be different across the cut-off. Data used are at the grade by year level. Fixed effects models include fixed effects for grades within schools.

Table A.17 Estimates of the effect of the experience components on test scores in 2009

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Discontinuity sample	All	+/- 1.5	+/- 1.0	+/- 0.5	All	+/- 1.5	+/- 1.5	+/- 0.5
1st-order polynomial			**Math**				**Language**	
Total experience	0.009	0.004	0.015	0.014	0.003	0.002	0.006	0.005
	(0.008)	(0.009)	(0.010)	(0.016)	(0.006)	(0.007)	(0.008)	(0.013)
Tenure	0.039***	0.037**	0.061***	0.060*	0.025**	0.025**	0.039**	0.019
	(0.014)	(0.015)	(0.018)	(0.030)	(0.011)	(0.012)	(0.018)	(0.029)
Experience at other schools	-0.009	-0.019	-0.018	-0.015	-0.007	-0.009	-0.011	0.010
	(0.011)	(0.012)	(0.012)	(0.023)	(0.008)	(0.009)	(0.011)	(0.020)
2nd-order polynomial								
Total experience	0.009	0.006	0.012	0.006	0.004	0.003	0.003	-0.008
	(0.009)	(0.009)	(0.010)	(0.019)	(0.006)	(0.007)	(0.008)	(0.013)
Tenure	0.039***	0.037**	0.057***	0.075*	0.026**	0.026**	0.032*	0.023
	(0.014)	(0.015)	(0.018)	(0.043)	(0.011)	(0.012)	(0.017)	(0.035)
Experience at other schools	-0.009	-0.018	-0.020*	-0.016	-0.008	-0.007	-0.015	0.001
	(0.011)	(0.013)	(0.012)	(0.022)	(0.008)	(0.009)	(0.011)	(0.018)
Observations	3,108	2,366	1,263	690	3,389	2,577	1,387	741

Notes: Test scores are regressed on total experience or on the two components (tenure, experience at other schools). All models control for age and gender. The three discontinuity samples (+/- 1.5; +/- 1.0; +/- 0.5) are based on standard deviations of the poverty score across the cut-off. Standard errors are adjusted for clustering at the school level.

Table A.18 Reduced form estimates of the effect of the CSCC program on student enrolment 2009-2013

	(1)	(2)	(3)	(4)	(5)	(6)
Discontinuity sample	+/- 1.5	+/- 1.0	+/- 0.5	+/- 1.5	+/- 1.0	+/- 0.5
School size		**All grades**			**Grades 1-3**	
1st-order polynomial	9.135***	9.209***	3.057	4.716***	5.940***	3.623
	(3.000)	(3.353)	(4.747)	(1.789)	(2.046)	(2.883)
2nd-order polynomial	6.598	-4.749	-10.076	4.422*	-2.045	-4.773
	(4.528)	(5.262)	(8.259)	(2.682)	(3.200)	(4.868)
Observations	2,549	1,930	1,100	2,551	1,932	1,101
Student background		**Lunch (0-4), all grades**			**Supper (0-4), all grades**	
1st-order polynomial	-0.114*	-0.221***	-0.149	-0.058	-0.139	-0.335*
	(0.069)	(0.084)	(0.121)	(0.112)	(0.139)	(0.194)
2nd-order polynomial	-0.164	-0.013	0.168	-0.234	-0.507**	-0.532*
	(0.106)	(0.130)	(0.188)	(0.174)	(0.216)	(0.290)
Observations	2,416	1,838	1,061	2,32	1,776	1,026
Schools	543	413	235	543	413	235

Notes: The dependent variables are school size or student background characteristics (lunch or supper at school). These dependent variables are regressed on program eligibility using the same specifications as in Table 2. Standard errors in models for lunch and supper participation adjusted for clustering at the school X year level.

Appendix A Primary education in Uruguay and the history of the CSCC program

The public education system in Uruguay has approximately 2,000 primary schools. Public schools are grouped in 5 main categories: Rural (*Escuelas Rurales*), Standard Urban (*Urbanas Comunes*), CSCC (*Contexto Socio Cultural Crítico*), Double Shift (*Tiempo Completo*) and Practice (*Habilitadas de Práctica y Práctica*). In Table A.19 we show the number of schools in each category since 1992. Rural Schools are very small schools located in the countryside. Although they are the majority of school facilities (56%), they cover a very small fraction of students (7.2% in the period under analysis). The typical rural schools has on average only 1.4 teachers in charge of the 6 grades of primary education, and has on average 18 students aged 6 to 12 years old. The majority of Uruguayan primary school students attend Standard Urban schools (Section 2). These Standard Urban schools could receive the extra resources from the program. The number of Standard Urban schools has steadily decreased, from 792 schools in 1992 to 400 in 2010. The same building facility of the Standard Urban school was used to implement one of the other types of schools that started to operate in that decade: CSCC schools, Double Shift and Practice.

The CSCC program started to operate in 1995. Between 1995 and 1999 it was called *Requerimiento Prioritario*; from that year until 2011 it was named *Contexto Socio Cultural Crítico*. Since 2011 the compensation program for poor schools is named A.PR.EN.D.E.R. (in Spanish means "to learn"), and is an acronym for *Atención Prioritaria en Entornos con Dificultades Estructurales Relativas*. The assignment criteria for the program have changed during the years. However, the goal of the program remained unchanged; compensating students with a disadvantaged family background.

In the year 1995 the first 155 schools started to participate in the CSCC Program (Table A.19). The majority (147) were schools that were already functioning as *Standard Urban* when they entered the program. In that first year of operation, schools were assigned based on school indicators of poor performance (i.e. grade retention) and characteristics of the houses in the neighbourhood were each school was located (based on information from census from the

National Institute of Statistics). This first allocation of schools didn't use direct measures of socioeconomic characteristics of the students attending the schools that were going to be treated.

Table A.19 School and students by type of primary school in Uruguay 1992-2013

Year	Type of School											
	Standard Urban		Rural		CSCC		Double Shift		Practice		TOTAL	
	Schools	Students	Schools	Students	Schools	Students	Schools	Students	Schools	Students	Schools	Students
1992	792	274,956	1,246	28,390	-	-	-	-	-	-	2,038	303,346
1993	787	271,990	1,237	27,236	-	-	39	4,106	-	-	2,063	303,332
1994	789	271,922	1,320	31,004	-	-	40	4,813	-	-	2,149	307,739
1995	651	223,658	1,307	30,881	155	51,291	46	6,552	-	-	2,159	312,382
1996	653	232,336	1,275	31,379	151	51,755	49	7,848	-	-	2,128	323,318
1997	636	231,138	1,240	27,514	149	54,220	58	9,418	-	-	2,083	322,290
1998	677	248,982	1,214	24,823	156	56,440	57	9,492	-	-	2,104	339,737
1999	507	173,639	1,183	23,066	273	107,819	66	11,875	-	-	2,029	316,399
2000	555	167,036	1,095	18,807	271	110,333	75	15,217	-	-	1,996	311,393
2001	718	236,610	1,094	18,686	106	43,540	84	18,869	90	41,183	2,092	358,888
2002	676	233,582	1,098	19,392	140	49,530	92	21,419	84	38,685	2,090	362,608
2003	631	214,804	1,089	19,985	151	55,412	95	22,451	114	51,892	2,080	364,544
2004	602	205,394	1,089	20,101	148	54,366	102	24,900	133	60,487	2,074	365,248
2005	592	200,035	1,092	20,282	150	54,345	104	25,160	135	60,296	2,073	360,118
2006	495	179,129	1,146	24,132	185	67,290	109	26,528	132	58,107	2,067	355,186
2007	463	158,537	1,143	23,534	221	86,166	111	26,256	126	53,547	2,064	348,040
2008	406	138,898	1,137	23,384	280	102,123	120	28,945	115	48,074	2,058	341,424
2009	395	132,817	1,142	23,486	285	101,438	132	31,359	114	46,881	2,068	335,981
2010	400	130,005	1,133	21,902	285	98,171	134	31,313	115	46,396	2,067	327,787
2011	383	121,577	1,132	21,136	271	96,476	157	34,937	117	45,849	2,060	319,975
2012	366	112,839	1,131	20,788	271	94,178	170	36,885	127	47,918	2,065	312,608
2013	329	104,226	1,107	19,429	265	90,543	188	40,400	129	48,060	2,018	302,658

Authors own calculations based on Monitor Educativo Educación Primaria (ANEP)

The program had an expansion between 1998 and 1999 and the program reached 273 schools. In the 1999 reallocation, schools entered or left the program based on three indicators: grade retention, insistences of students in 1st grade, and the percentage of students in 6th grade whose mothers had primary education as the highest level of formal education (with data from 1996) (ANEP 2005).

In 2001 the program was reduced to 106 schools. In the 2002 reallocation, the program had a 32% net increase in the number of schools participating, when 85 new schools joined the program (80% increase considering the schools in the previous year), and 51 schools were dropped (48% of the schools in the previous year). The criteria used for the assignment to the CSCC program were based exclusively on socioeconomic variables of students. The variables used by the central authority were the percentage of children: (i) with unemployed household heads (or doing very informal jobs); (ii) whose mothers didn't achieve primary education; (iii) that were allowed to receive free lunch at school; and (iv) that lived in overcrowded houses. Each state in the country had a fixed slot of schools that were going to receive the extra resources, based on the number of students in the state and the percentage of children between 4 and 12 years old in the poorest quintile of the distribution of income (ANEP 2005)

From 2003 to 2005 the number of schools participating was around 150. In 2005 a new categorization of schools was made, and in the following years new schools entered the program, and others left because families living in the neighbourhoods in which they were located improved their socioeconomic status. The increase in the coverage rate of the program that started in 2006 ended in 2009, when 285 schools were receiving the extra resources, and the number of schools participating reached its maximum in 10 years. The 2005 re-categorization (schools moving in and out of the program) used indicators of human capital (level of education of mothers or, in her absence, the adult in charge of the children), socioeconomic level (unmet basic needs index), and an index of social integration. All this information was summarized (using factor analysis for data reduction) in a single measure that described the context of each school (cfr ANEP 2005). All the schools were then ranked according to this unique index, which is the forcing variable that we will use to perform the RD analysis. We focus our analysis on this period since the assignment to the program was fully transparent for these years. The eligibility rules changed again in 2011. Since then the assignment to the program was no longer based on the threshold of the poverty index from 2005.

Appendix B Identification issues

The main assumption in the regression discontinuity model is that all observed and unobserved factors should behave smoothly around the cutoff. To test this assumption we first look at the density of the forcing variable at the cutoff to investigate whether schools might have manipulated their assignment to the program. It should be noted that this type of manipulation is not very likely as the assignment to the program is completely determined by the central authority (see Section 2). Figure A.2 shows the density of the forcing variable across the cutoff based on the methods by Cattaneo, Jansson and Ma (2017). If schools would have manipulated their eligibility for treatment we would expect a larger density at the right side of the cutoff. However, we don't observe this in the data. If anything, the density appears to be slightly, but statistically insignificant, higher at the left side of the cutoff. Both the conventional test as the robust bias-corrected test yield statistically insignificant results. Hence, these tests don't indicate that the assignment to the program has been manipulated.

As a second test we perform balancing tests of covariates and outcomes in the baseline year (2005). Figure A.3 shows the results of these tests for the outcomes in 2005 and several covariates. Table A.4 in the appendix shows the balancing tests using three discontinuity samples around the cutoff. These tests suggest that schools on both sides of the cut-off were very similar on teacher and student outcomes in 2005; we only observe a difference in school size.

A further concern with our empirical analysis is that the program already exists since 1995. Differences in program participation in the years before the redesigning of the program might confound the estimates. If schools on either side of the cutoff received more resources from earlier program participation this might bias the results. To investigate this issue we have estimated Equation (1) for each year since 1995 using as dependent variable 'participation in the CSCC-program'. For the whole period since 1995 we find that there were no differences in program participation at the cutoff that might confound our estimates. These results are shown in the next sections (Figure 2 and Figure A.5).

Another concern with our empirical approach is that there might have been changes in the composition of students in schools across. Although the CSCC-program explicitly focused on changes in the teaching staff it might also have affected the targeted schools along other margins.

For instance, the hiring of more experienced teachers might have made these schools more attractive for students and their parents. An increase in school size might have reduced student performance through larger classes. Moreover, changes in the socioeconomic background of students might also have an impact on student performance. It should be noted that changes in the enrolment of students are limited by the application of fixed catchment areas in primary education in Uruguay. We investigate whether the CSCC-program had an impact on school size and on the socioeconomic background of students for the most relevant period of the program. We estimate the main models from the previous sections, and use as dependent variables school size and two indicators of socioeconomic background of students; proportions of students in need of lunch or supper at school. The need of lunch or supper at school indicates poverty at home; both variables have five categories: 0%, 0-24%, 25-49 %, 50-74 %, 75-100%, and are measured yearly at the school-level.

The top panel of Table A.18 shows the impact of being eligible for the program on school size. Note that in our specification we control for school size in 2005, which implies that the estimates measure the increase of schools size since the redesigning of the program. The estimates from models that use a linear specification of the forcing variable suggest a small increase in school size. The estimated effects of 3 to 9 students imply an increase of class size with 0.25 to 0.75 students as schools on average have 12 groups. The estimates that use a quadratic specification don't indicate an increase of schools. The bottom panel shows the effects on the proportions of students that need lunch or supper at school. We don't observe a clear pattern for the first indicator (lunch). The estimates for the second indicator suggest that the proportion of students in need of supper at school is somewhat smaller in schools that were eligible for the program. This suggests that the poverty rate of students in schools at the right side of the cutoff is somewhat lower than in schools at the left side of the cut-off. Hence, we don't find that the proportion of disadvantaged students has increased due to the program. These analysis on changes in enrolment of students don't provide evidence that the program affected the targeted schools along other margins in such a way that it could explain the modest results of the program.

Robustness analysis

For our main analyses we estimate models that include linear and quadratic specifications of the forcing variable. We investigate the robustness of the estimates to different specifications of the forcing variable. In particular, we use a cubic specification of the forcing variable and also use local polynomial Regression Discontinuity (RD) point estimators (first order and second order polynomials) with robust bias−corrected confidence intervals as developed in Calonico et al. (2014). In our main models we control for grade fixed effects, year fixed effects and, due to the balancing test, for school size (using a quadratic specification[25]). We also test the robustness of the results to including pre-treatment outcomes from the period before the redesigning of the program in 2005 and to including additional indicators of family background ('mothers with primary education or less', 'students with unmet basic needs'). As RD empirical results are often sensitive to the choice of the bandwidth, we show the results for different bandwidths. For our main estimates we use three discontinuity samples around the cut-off value of the forcing variable; schools within the ranges of 1.5, 1.0 and 0.5 standard deviations of the poverty score across the cut-off. Data-driven bandwidth selectors, as proposed in Calonico et al (2014) and Cattaneo et al. (2018), yield optimal bandwidths that correspond with the range of our discontinuity samples. The optimal bandwidth varies and depends on the bandwidth selector, the type of model (sharp or fuzzy regression discontinuity model), the model specification and the sample size.

As we have data for multiple years since 2005 we pool the data for the relevant years (and include year dummies) to improve the precision of our estimates. An advantage of having multiple years of data is that our estimates will be less sensitive to confounding factors related with implementation issues in the initial years. Schools and teachers might need some time to adjust their decision to the new rules of the program. Hendricks (2014) notes that teacher pay may have a direct effect on the quality of the school but it is not clear when there will be an effect on student outcomes. With our data we can investigate the effect of the program up till 2013[26].

[25] A linear specification yields similar results.

[26] It should be noted that we are not estimating 'long term effects' of the program in models in which year t outcomes are regressed on program participation in year t-1 or year t-2.

Systematic differences in missing values between schools might bias the results. Our data are obtained from administrative registries and collected for administrative reasons which mitigates this concern. To probe the randomness of the missing values we have estimated our main regression models using a dummy for missing outcome variable as dependent variable. These analyses don’t yield concerns about systematic differences in missing outcomes which might bias our results.