Does attending a selective secondary school improve student performance? Evidence from the *Bicentenario* schools in Chile

Pablo Araya & Francisca Dussaillant

To cite this article: Pablo Araya & Francisca Dussaillant (2020) Does attending a selective secondary school improve student performance? Evidence from the *Bicentenario* schools in Chile, School Effectiveness and School Improvement, 31:3, 426-444, DOI: 10.1080/09243453.2019.1697299

To link to this article: https://doi.org/10.1080/09243453.2019.1697299

School Effectiveness and School Improvement **An International Journal of Research, Policy and Practice**

ISSN: (Print) (Online) Journal homepage: https://www.tandfonline.com/loi/nses20

Published online: 03 Dec 2019.

SCHOOL EFFECTIVENESS AND SCHOOL IMPROVEMENT 2020, VOL. 31, NO.
3, 426-444 https://doi.org/10.1080/09243453.2019.1697299

Does attending a selective secondary school improve student performance? Evidence from the *Bicentenario* schools in Chile

Pablo Araya[a] and Francisca Dussaillant [b]

[a]Ministerio de Educación, Santiago, Chile; [b]School of Government, Universidad del Desarrollo, Santiago, Chile

November 2019

ABSTRACT We investigate the effect of attending selective secondary schools belonging to the *Bicentenario* schools program in Chile, a free education option that is intended to give vulnerable students everywhere in Chile a unique educational opportunity, in the period from 2011 to 2014. By using propensity score methods, we find that attending a *Bicentenario* school improves students' performance by a range of 0.35 to 0.23 *SD* in language and 0.5 to 0.35 *SD* in math. Also, we show that a proportion of this effect is due to the outstanding performance of the new *Bicentenario* schools as opposed to the restructured version of the program. We prove the robustness of the previous results through falsification, changes-in-changes, and a more demanding control group. This paper adds new evidence to analyze the effect of selective schools in developing countries like Chile, where the quality of public education is low.

ARTICLE HISTORY Received 28 August 2018 Accepted 15

KEYWORDS Education policy; selective schools; effective schools; public schools; program evaluation

Introduction

Since the year 2000, Latin American countries have made great efforts to improve the quality of their education. Some facts illustrate this trend, for instance, the increase in the spending on public education rising from an average of 4.5% of gross domestic product to 5.2%; the high rate of access to primary education (94%); better access to early education (increasing from 56% to 66%); and a far-from-modest access to secondary education, at 72%. Nonetheless, these figures worsen once socioeconomic and demo- graphic differences are included that affect the most disadvantaged students from poorer families (United Nations Educational, Scientific and Cultural Organization [UNESCO], 2015). Despite the improvements, one of the

challenges remaining for Latin American countries is to improve the quality of education to achieve better learning out- comes and to benefit more disadvantaged groups and families.

Among Latin American countries, the Chilean educational system has shown improve- ments in international achievement tests (Organisation for Economic Co-operation and Development [OECD], 2014; UNESCO, 2015), but it is still far below the average of devel- oped nations. The Chilean P–12 school system (primary and secondary schools) relies heavily on a version of school vouchers introduced by the military government in 1981 that allowed the entry of new educational providers (private subsidized schools) to

CONTACT Francisca Dussaillant frandussa@gmail.com

compete with public schools administered by local municipalities. After three decades, private subsidized schools represent 56% of the national enrollment, while public school enrollment has decreased from 78% in 1981 to 38.5% in 2016. Evidence shows that public schools are still attended by the most at-risk families because the mass migration to private subsidized schools was disproportionately composed of the wealth- iest families in the public sector. This occurred because many of the new private subsi- dized alternatives opted for *a shared-financing* modality, available from 1993 onward, where they can charge parents moderate fees without losing most of the state subsidy (Hsieh & Urquiola, 2006; Lara, Mizala, & Repetto, 2011).

The Chilean system is characterized by huge outcome differences between the govern- ment-subsidized (private and public) schools and their non-subsidized private (paid) counterparts. Private non-subsidized schools account for 7% of the national enrollment, but their students are disproportionately represented later in the selective universities.

In 2014, the government launched a major reform to increase the quality of, and reduce the segregation within, the education system. This reform included different initiatives: an increase in government subsidies; the creation of new, centralized public agencies to manage public schools; the prohibition against private subsidized schools charging parents a fee; the elimination of the for-profit private subsidized schools; and the ban on selecting stu- dents in any grade by academic criteria since it could be construed as discrimination.

This last initiative gave rise to a huge debate considering that there is no conclusive evidence that the Chilean system was too selective and that there are pros and cons in the academic literature about selectivity in secondary schools (Rivera & Guevara, 2017). In fact, student records of academic performance and/or recommendations of feeder schools are considered in the admission process in a manner similar to that of OECD countries (OECD, 2013). There is evidence supporting that selecting students can be a means to achieve better results for all students given the benefit of attending a less dis- perse class that allows teachers to focus their teaching strategies (Fontaine & Urzúa, 2014). Accordingly, academic selection is a reasonable option in contexts where effective schools for the poor are scarce. This is the case in Latin American countries.

The ban on academic selection directly affected the *Emblemáticos* (Emblematic) schools, a small group of public secondary schools (about 20) that have historically yielded academic results similar to those of private non-subsidized schools. Access to Emblematic schools has always depended on aptitude test scores and previous school grades. Emblematic schools are located mainly in the capital and have a good reputation because a significant number of their students are admitted to the most selective Chilean universities. For instance, in 2015, only 18% of the students at the two most selective uni- versities in Chile came from public schools. Three out of four of these students had studied at an Emblematic school (Eyzaguirre, 2016). Emblematic school alumni have gained access to the political, cultural, and economic elite of the country. In fact, from a historical per- spective, Emblematic schools have been perceived as one of the most important means of social mobility for the lower middle classes, and a mechanism that helps renovate and diversify the elite (Fontaine & Urzúa, 2014). Emblematic schools represent the ideal of a meritocratic public education, one

of the very few means of social mobility in the country, since most public schools do not teach the minimum contents evaluated by the university entrance test (Koljatic & Silva, 2010). This leads to a low share of public-school students being admitted to those selective universities.

In 2010, before the recent wave of anti-selection reforms, a new program was launched, intended to replicate the Emblematic model in 60 schools throughout the country. It was called the *Bicentenario* (Bicentennial) Program. The program was in response to a per- ceived lack of access to adequate learning that was severely limiting the opportunities of high-achieving students nationwide. A few years after the implementation of the program, national test scores showed that Bicentennial schools were outperforming public and private voucher schools and, what is most remarkable, were obtaining results almost identical to private paid schools, even in terms of access to selective univer- sities (Rivera & Guevara, 2017). This effect seems to continue, even after the reform that limited student selection.

Therefore, our research question involves the effect of the Bicentennial program on the learning outcomes of the students that attended them. This paper reports our estimates of the impact of attending schools that joined the Bicentennial schools program by studying the eighth-grade standardized math and language test scores for 2013, 2014, and 2015.

The next sections are organized as follows. After the literature review, we describe the main features of the Bicentennial schools program. Then we provide details on the data and methodology used in this study. We show our main results in two steps: (1) We provide propensity-score estimates as a reference point, and (2) we provide the results of different strategies to check the robustness of our estimates. Concluding remarks and policy implications are offered at the end.

Literature review

The literature on academic projects that include selection in secondary schools is under debate. Research does not identify a clear pattern in the effect of selective schools on aca- demic achievement. In addition, research has the challenge to adopt the more suitable and available methods to deal with the difficulty of tracking.

Evidence for developed countries tends to suggest that selective schools produce no significant gains in terms of academic achievement. One of the cases most studied in developed countries is the banning of grammar schools in the United Kingdom during the 1960s and the promotion of a comprehensive school system. By using matching and instrumental variable methods, Galindo-Rueda and Vignoles (2005) found suggestive evidence that more capable students earn higher test scores in a selective system, so a comprehensive system would have a negative impact on their performance. However, Manning and Pischke (2006) suggested that the results of this reform should be inter- preted carefully because banning grammar schools was not a random process throughout the country, and it still could involve selection. In fact, they confirmed this problem, showing that in some areas, results do not differ pre-and post-treatment, even though the degree of selectivity in the area does. Further research has used proper and more robust methods. For instance, Clark (2010) used regression discontinuity to assess the effect of the selective system, finding modest and significant short- and long-term effects on both test scores and college enrollment. In the same way, Clark and Del Bono (2016) found positive long-term effects on college completion, but no effect on most of the labor-market outcomes. For the case of the United States, in their study of the most selective high schools in Boston and New York, Abdulkadiroğlu, Angrist, and Pathak (2014) found no significant effect on admitted applicants near the admission

process cut-off (marginal students), which would also imply that peer effects are not strong. In the same way, Dobbie and Fryer (2014) found little impact on college enroll- ment, college graduation, or college quality.

Evidence from developing countries is more compelling in suggesting that selective secondary schools may have a positive effect on test scores and college enrollment. Pop-Eleches and Urquiola (2013) used more than 2,000 regression discontinuities of Ruma- nian schools to find that the marginal students admitted increased their graduation test scores by 0.1 *SD*, although those students struggled to adapt to a more competitive regime and received less parental support. Jackson (2010) found that students who attended selective schools in Trinidad and Tobago were more likely to pass national exams and, moreover, that the school effect would not be driven significantly by peer composition (Jackson, 2013). Similarly, Dustan (2010) found that a selective group of schools in Mexico improved the test scores of students near the admission cut-off by 0.19 *SD*.

There is still little evidence in the case of Chile. Using propensity score methods, Allende (2015) assessed the effect of attending one of the 17 Emblematic schools on 10th-grade test scores to estimate the average treatment effect on the treated (ATET). The study found that attending an Emblematic school was associated with better test scores, and the mag- nitude of that effect ranged from 0.2 to 0.3 *SD*. However, the author suggests that this effect might only be reflecting positive peer effects.

Using similar methods, Henríquez, Lara, Mizala, and Repetto (2012) estimated the average treatment effect (ATE) of attending a private voucher school belonging to an edu- cational network that has proven to be effective at teaching at-risk students. The study's results show that the effect (ATE) of attending fourth grade at a school in the network is higher than 0.5 *SD* when compared with public schools, and no lower than 0.25 *SD* when compared with other private voucher schools. However, the authors argue that unob- served factors, like parent motivation, given the good reputation of these schools, may be a source of overestimation of the effect.

Bucarey, Jorquera, Muñoz, and Urzúa (2014) identified the effect of attending the most traditional and selective Emblematic school in Chile. This study found an effect of 0.25 *SD* on the Chilean higher education admissions test, which may potentially make a difference in university applications. Interestingly, these results suggest that the second-best option for students not making the cut-off in that school's selection process (probably another Emblematic school) is not adding value to students; or that peer effects are strong there. Using the same method, Manríquez (2016) did not find an effect of attending a girls-only Emblematic school in Chile on 10th-grade test scores.

Finally, Alcalde (2014) evaluated the effectiveness of the Bicentennial schools based on a model of student effort choice (scholastic effort), concluding that the restructured or existing Bicentennial schools increased their mean productivity once they entered the program, but the effect is heterogeneous. She also noted that the new Bicentennial schools displayed outstanding productivity rates; three of them were even among the 20 most productive schools in the country.

In conclusion, selective secondary schools in developing countries seem to be more effective than in developed countries. One possible explanation is the huge quality breach between the best schools (that use selection to manage their excess demand) and most schools in the system.

The Bicentennial schools program

The Bicentennial schools program is aimed at promoting the equal-opportunity principle, targeting talented students from disadvantaged families by giving them the chance to enroll in a school that increases their likelihood of access to selective higher education. The program is nationwide, in all 15 regions and in 59 out of 350 municipalities (*comunas*)[1] across the country.[2] The schools joined the program in two stages, 30 in 2011 and 30 in 2012. The program's intervention focused on the seventh grade in 2011, and on the seventh and eighth grades in 2012 and 2013. From 2014 onward, the schools

no longer received pedagogical support due to the change in government administration.

We can classify Bicentennial schools as New, Restructured, or Expanded, depending on their origins. New schools (20 in total) correspond to completely new academic-oriented schools that were opened to the public with educational projects aligned with the purpose of the program from the beginning. As a matter of fact, the principals of these schools were able to hire teachers under competitive processes. In some cases, teachers were offered financial incentives or higher salaries conditional on the academic achievement of their students. Restructured schools (38) existed before the program was launched, but transformed their educational project to meet the program's targets and did not make significant changes in their professional teams due to bureaucratic obstacles to remove less effective teachers. Expanded schools (2) were schools that already met the targets of the program, so they were merely asked to join and to increase their capacity (these were not included in our analyses). Table 1 describes the main features of the Bicen- tennial program.

Implementation facts

The Ministry of Education (MINEDUC) formed a team to monitor, support, and advise the schools. At the beginning of the academic year, all Bicentennial schools made an initial assessment of seventh graders' math and language to obtain diagnostic information. The diagnostic test showed that the students knew barely 50% of the contents for fourth and fifth grade, so the schools added a new remedial unit to the curriculum for

Table 1. Bicentennial schools program description. Who could participate?

All subsidized schools. Public and private voucher schools could apply. However, public schools were the focus of the program. In fact, the final composition was only five private voucher schools and 55 public schools. The projects were presented by the owner of the voucher schools or by the mayor, in the case of public schools. Size Schools should open at least 80 places in seventh grade. Benefits New schools could request funding of as much as US$1.8 million to invest in infrastructure and educational technologies or equipment; restructured and expanded schools up to US$0.9 million. The program did not provide monetary incentives to teachers. Selection decision rules

The educational project (the more academic oriented, the better); specific initiatives to attract disadvantaged students; geographic criteria, teachers and principal qualifications. All factors equally weighted. Targets Schools should be among the top 10% in the SIMCE* test results and the top 5% in the PSU test results. Note: *SIMCE, Sistema de Medición de la Calidad de la Educación (Education Quality Measurement System), is the national standardized test administered every year in fourth grade, and in either eighth grade or 10th grade, that is, in year t SIMCE is given to fourth and eighth graders and, in year t + 1, to fourth and 10th graders. Source: Authors, based on Carrasco, Gutierrez, Bogolasky, Rivero, and Zarhi (2014).

that year (Bravo, Lavín, Ruiz de Viñaspre, & Wilkins, 2016). If the class average did not exceed 80%, school principals were advised to repeat the test until the target was reached. Additionally, the team suggested a change in the order of the math curriculum for seventh grade; the idea was to start with easier topics, such as integers, and not fractions, facilitating the math learning process and taking into account the results of the diagnostic test. Despite these proposed changes, the Bicentennial schools teach the entire national curriculum required for all Chilean schools that is measured by the national test, called SIMCE. However, they offered it in a more flexible way, prioritizing some critical contents, and reviewing contents from previous years when necessary. The team deployed successful strategies of math and language teaching and learning scripts and principal and teacher training to improve results. A similar strategy can be found in a project implemented in the municipality of Calama, Chile (Schiefelbein, Leiva, & Schiefelbein, 2004).

Contrary to expectations, Bicentennial schools were not oversubscribed. Factors such as geographical location, city population, and negative parental view of public education could explain this problem (Carrasco et al., 2014, pp. 64–65). However, admission rules were enough to attract high-achieving students. Most of the Bicentennial schools required a minimum (approximate) grade point average (GPA) of 6.0 in fifth and sixth

grades to apply, plus admission tests. In the 1.0 to 7.0 Chilean grading scale, where a 4 is the minimum passing grade and the average in eighth grade is slightly below 5.6 (González & Johnson, 2018), a GPA of 6.0 can be considered "good." Imposing selection requirements was later deemed illegal under the 2015 education reform. But the cohorts under study were admitted to the programs under the previous legal framework that allowed selection for seventh grade and higher.

Methods and data

Our main methodological challenge was to overcome several sources of selection bias, such as the non-random assignment of students. In fact, selection is reinforced by the program's focus on high-achieving students. Literature has dealt with the selection problem mainly by using the regression discontinuity method (the discontinuity induced by the admission rule), but this was not possible in our case because many Bicentennial schools were not oversub- scribed (and, therefore, they did not have the chance to explicitly select their students). However, self-selection based on student motivation or parental valuation of education, both unobservable, may be a factor given that applying to a Bicentennial school is voluntary. We used propensity score methods to address this issue (see, e.g., Adelson, 2013). Specifically, we used the inverse probability weighting and regression adjustment to esti- mate the treatment effects of attending a Bicentennial school, and we tested whether our results were robust by implementing a falsification test using fourth-grade test scores as outcomes and by estimating an alternative model according to Athey and Imbens' (2006) changes-in-changes (CIC). Finally, we repeated the estimations with a different and more restricted control group, the Emblematic schools. This exercise was valuable because it is reasonable to think that unobservable inputs, such as student motivation and parent com- mitment, are similar across these educational projects as they are both performance oriented and share similar educational purposes. Moreover, Emblematic schools have historically been highly selective (more than Bicentennial schools), so we can be sure that any advantage of the latter over the former is not the result of peer effects.

Data

We used data from the SIMCE standardized national test. This dataset includes math and language test scores and a parent questionnaire that is administered in each test round. The questionnaire provides information about families' socioeconomic characteristics and other relevant information, such as the perception that the parents have of their chil- dren and of the school they attend. The data we used was a panel; we had information on students in fourth grade (the "base" year) and 4 years later (eighth grade), 1 year after some of these students started attending a Bicentennial school (we only evaluated Bicentennial Schools that started at seventh grade). The dataset was restricted to all those observations for which there was complete socioeconomic information at least in the "base" year to allow for a pre-treatment control and had valid test scores both in fourth and in eighth grade, in at least one discipline (math or language).[3] We also matched this dataset to the final grade point average earned by the students in the fifth and sixth grades, which was made available by the Ministry of Education for research purposes (both SIMCE and GPA data are nameless, but the Ministry uses a masked student identifier that is constant across its databases, allowing us to match GPAs and SIMCE scores student by student).

Propensity score methods

The propensity score is defined as the probability of being treated based on a set of obser- vable

characteristics. It relies on the creation of a control group (ex-post) from a pool of individuals who could have received the treatment based on their observable character- istics. If this control group properly emulates the counterfactual, it would be possible to estimate how Bicentennial students would have performed if there were no program.

We defined our control group following the procedure of Allende (2015) in his study on Emblematic schools. Given that 47 out of 60 Bicentennial schools receive their students in seventh grade, these students attended a different school in fourth grade. Our data allowed us to retrace them and identify their fourth-grade classmates, so we constructed our control group from these former classmates. Ten Bicentennial schools that do provide instruction for lower grades were excluded from our analysis. The control group was further restricted to students who had a final GPA above 5.0 in fifth and sixth grade and were categorized, according to the SIMCE test, as "elementary"[4] in math or language in fourth grade. Therefore, the counterfactual comes from individuals in the control group who meet the requirements to apply to a Bicentennial school and who are similar to their treated counterparts in terms of several relevant observables described below.

Propensity score methods rely on strong assumptions. The assumption of conditional independence or *unconfoundedness* entails no unobserved variables associated with either the potential outcomes or the treatment (Imbens & Wooldridge, 2009, p. 26). It implies that, after controlling for the set of covariates, the assignment to the treatment is as good as random. Second, there is the assumption of overlap or common support. This assumption says that it is possible to find individuals with the same propensity score both in the treatment and in the control group.

The most popular propensity score estimation method is propensity score matching (PSM), roughly consisting of comparing the average outcome of individuals that were and were not treated but who had similar propensity scores (Heinrich, Maffioli, &

Vázquez, 2010). Despite its popularity, this strategy has some limitations in terms of efficiency (Khandker, Koolwal, & Samad, 2009). We therefore decided to use two additional propensity score estimation methods that address those limitations (Lara et al., 2011). Inverse probability weighting (IPW) is a weighted regression model that uses the inverse of the propensity scores (PS) as weights that, theoretically, counteract the effect of selection (Imbens & Wooldridge, 2009). Inverse-probability-weighted regression adjust- ment (Double-Robust) is an improved version of IPW (Cerulli, 2015, p.133).

Propensity scores were constructed for all our models via logistic regression, using a set of variables measured before the beginning of the program. Some of these were measured at the individual level and others at the school level. Among the former are gender, the student's GPA in fifth and sixth grades, the student's SIMCE language and math scores in fourth grade, the mother's education level (in years), the father's education level (in years), monthly household income (in thousands of Chilean pesos[5]), a dummy for whether the student attended pre-school, a dummy variable indicating whether the student repeated a grade, a dummy variable for whether the family expected the student to reach higher education, and the number of books at home. At the school level, we included the type of school the student attended (private, private voucher, or public) in sixth grade, a dummy for indicating whether the school is located in a rural zone, and total school enrollment. We included an interaction between the average of fifth and sixth grades and type of school to capture previous schools' quality and academic demand differences. At this stage, estimations were giving us a way to answer the research question of *What is the probability of student types attending a Bicentennial school?*

We estimated the propensity score[6] and imposed the common support restriction fol- lowing the recommendations of Garrido et al. (2014). We also assessed the quality of our control group. For instance, the degree of overlap between treated and untreated was assessed visually (this figure is available from the

authors upon request), and the imbal- ance was checked across covariates following the recommendations of Garrido et al. (2014, p. 1708). Standardized differences no higher than one fourth of a standard differ- ence were considered to be an acceptable measure of group balance (Imbens & Wool- dridge, 2009).

We provided estimates of the average treatment effect on the treated (ATET) by type of Bicentennial school (new and restructured) to see if there were different effects between both versions of the project. The ATET allowed us to answer the research question of *What is the score improvement a Bicentennial student achieves as compared to the score the same student would have earned at another (non-Bicentennial) school?* Depending on the group used for comparison, the question: *What is the score improvement a Bicentennial student achieves as compared to the score the same student would have earned at an Emblematic school?* is also answered.

Robustness checks

The key assumption for matching and weighting methods is unconfoundedness, namely, that selection depends only on observables. Imbens and Wooldridge (2009) suggest a fal- sification test to examine the reliability of this strong assumption. Falsification requires the use of a new outcome variable, similar to the outcome of interest, but which should not affect the treatment. In our case, pre-treatment test scores in fourth grade meet these

requirements. If the unconfoundedness assumption is true, the treatment should have no effect on SIMCE fourth-grade test scores given that we focus on the Bicentennial schools that start at the seventh grade.

We also exploited our data and created a panel to implement a changes-in-changes (CIC) analysis that utilized variation across treatment groups and time periods. This model, initially proposed by Athey and Imbens (2006), is a generalization of the traditional differences-in-differences method that does not impose the common trend assumption (namely, observations in the treatment and control group would behave in the same pattern in absence of the treatment). CIC analysis is valuable because it estimates the dis- tribution of unobservables under two assumptions: A higher level of unobservables implies higher outcomes, and unobservables remain constant over time. To implement CIC estimates, we followed Melly and Santangelo (2015), who suggest a flexible, semipara- metric estimator based on quantile regressions to identify the effect along the entire dis- tribution. For these estimations, we included the same covariates as those used when estimating through propensity score models, and we added covariates that captured changes in peer composition, such as income and parents' level of education.

Finally, we repeated our estimations using a more demanding control group: the Emblematic schools. As we mentioned in the Introduction, the Bicentennial program was inspired by the Emblematic school model, so it is reasonable to think that unobserva- ble variables like student motivation and parent commitment would be similar for Emble- matic schools and this new control group. Moreover, since the Emblematic schools are located in the capital city (the most populated city in the country), while the Bicentennial schools are scattered throughout Chile, sometimes in rural areas, Emblematic schools are better positioned to conduct a stringent selection process. Emblematic students applied to their schools with fifth- and sixth-grade averages that were almost the same as those of Bicentennial students. On the other hand, Emblematic students outscored their Bicenten- nial counterparts by an average of 20 fourth-grade SIMCE points (almost 0.5 *SD*). Moreover, average household income and years of parent education are higher in Emblematic schools. This is important because any advantage of the Bicentennial schools over the Emblematic schools could hardly be interpreted to mean that the Bicentennial schools are more successful at "cream skimming" or achieving, through selection, a "better" peer mix in order to exploit positive peer effects.

Main results

Probability of attending a Bicentennial school

Table 2 shows the results of the logistic regression model to predict the probability of attending a Bicentennial school for our three different cohorts of students. To facilitate interpretation, we provided the marginal effects. Results show positive marginal effects of gender (male), particularly significant for the second cohort (2014). These results are in line with evidence that indicates that females tend to avoid the more competitive environments like those found in elite education (Koch, Nafziger, & Nielsen, 2015). Fifth and sixth grades (GPAs) are significant in explaining enrollment only for public schools (in 2013 and 2014) and voucher schools (but only in 2014). Also, contrary to our expec- tations, previous test scores were not good at predicting enrollment. For instance, the

Table 2. Probability of attending a Bicentennial school. Variables 2013 2014 2015 Gender (1 if student is male, 0 if female) 0.0048086
(0.0040743)
0.0114906 ** (0.0037164)
0.0034921 (0.0033537) StudentTs GPA in fifth and sixth grade 0.0031961
(0.0052058)
0.0029087 (0.0040119)
0.0054942 (0.004585) SIMCE fourth-grade math scores 0. 0001125 (0.0000619)
0.0001418 ** (0.000055)
0.0000748 (0.0000505) SIMCE fourth-grade language scores −0.0004179 ***
(0.000066)
−0.0003035 ***
(0.0000636)
−0.0001848 ***
(0.0000525) Repeated grade (1 if YES, 0 if NO) −0. 0067452 (0.0150128)
−0.0231436
(0.0144267)
−0.0268864 * (0.0116973) Attended preschool (1 if YES, 0 if NO) 0.0028678
(0.0050208)
0.0070255 (0.0048608)
0.0195509 *** (0.0051064) StudentsT mean family income in fourth grade −0.0000353 ***
(0.00000835)
−0.0000203 ** (0.00000701)
−0.0000239 *** (0.00000609) FatherTs education 0.0018415 **
(0.000763)
0.0005231 (0.0006868)
0.0021609 *** (0.0006092) MotherTs education 0.0034967 ***
(0.0008296)
0.0007116 (0.0007374)
0.0353301 *** (0.0068213) Expectations: University (1 if YES, 0 if NO) 0.0401025 **
(0.0144066)
0.0386552 * (0.0153054)
0.0353237 *** (0.0068217) Expectations: Technical education (1 if YES, 0 if NO) 0.0036729
(0.0162196)
0.0124513 (0.0172669)
0.0098759 (0.0080935) Number of books at home (1 if books > 50, 0 if none) 0.0068242
(0.0051686)
0.015277 *** (0.0044984)
0.0053142 (0.0040981) Public school (1 if public, 0 if not) −0.190586 ***
(0.0570529)
−0.0989666

(0.0520601)

−0.0902198 * (0.0460619) Private (non-voucher) school (1 if private, 0 if not) −0.5587032

(0.3172543)

0. 5133294 * (0.2489128)

0.4011762 (0.2788605) Rural (1 if rural, 0 if urban) 0.0416621 ***

(0.0070617)

0. 0179256 ** (0.0061652)

-

Sixth-grade school enrollment - −0.0000568 ***

(0.00000472)

−0.0000521*** (0.00000377) Interaction fifth/sixth grades x public school 0. 0116699 *

(0.0052384)

0.0111845 ** (0.004048)

0.0072465 (0.0046106) Interaction fifth/sixth grades x public vouchers 0.0074082

(0.0052417)

0. 0088484 * (0.0040486)

0.004783 (0.0046067) $N_{LR\ Chi^2\ (17)}$ *Prob > Chi² Pseudo* R^2 *Model correctly classified*

19,890 1,364.00 0.0000 0.1111 90.70%

23,255 1,787.56 0.0000 0.1286 91.15%

29,461 2.129.06 0.0000 0.1192 90.98% Note: Standard errors in parentheses. *$p < 0.05$. **$p < 0.01$. ***$p < 0.001$.

marginal effects for SIMCE fourth-grade math scores were small and significant only at the 90% level in the 2013 cohort and at the 99% level in 2014, but they were not significant in the third one (2015). The probability of attending a Bicentennial school decreased (slightly) with higher SIMCE fourth-grade language scores. These figures reinforce the idea that the program serves a group of good students who are better in terms of school grades, but who are not necessarily superior in terms of overall performance in national test scores.

Parents' educational level is positively correlated with attending a Bicentennial school. The marginal effect is higher for mothers than for fathers. The probability of being admitted to a Bicentennial school is about 4% higher for students whose parents think that their children will go to university. These results are consistent with the idea that par- ental expectations and educational level are associated with educational decisions. Inter- estingly, students from lower income families have a higher probability of being admitted to a Bicentennial school, which indicates that the program's focalization criteria are met: Students from poorer families are the ones receiving the treatment.

At the school level, results show that coming from a rural school increases the prob- ability of attending a Bicentennial school. This may be true because many rural schools only provide instruction through sixth grade. These students are forced to find an edu- cational alternative for their last years of primary education (seventh and eighth grade) and secondary education, and the Bicentennial schools offer them a real alternative. Alter- natively, in an urban setting, it is more common to find schools that offer instruction for the complete primary and secondary cycles.

Average treatment effect on the treated (ATET)

Table 3 presents the ATET estimates on attending a Bicentennial school based on the SIMCE eighth-grade math and language test scores for our three different cohorts. Four facts can be highlighted from the results. First, student test scores seem to improve signifi- cantly when attending a Bicentennial school. For language, the ATET ranges from 0.42 to 0.22 *SD*; and for math, from 0.59 to 0.30 *SD*. From the viewpoint of Murnane and Willett (2010), most of the identified effects should be considered large since they exceed their established threshold of 0.3 *SD*. Second, IPW estimates are slightly higher than Double- Robust (DR) ones. The DR method captures changes in peer composition (in terms of family income and parents' educational level), which may explain this difference. The third fact is a decreasing pattern in the magnitude of the Bicentennial school effect, but with a sharper decline in 2015. We are not

able to fully explain this trend, but one of the hypotheses that can be studied in future research is whether it is related to changes in the program's leadership, to the school being more motivated about the program in the initial stages, or to changes in admission policies. We also suggest that

Table 3. ATET estimation bicentennial school.

ATET - Language eighth grade ATET - Math eighth grade Method 2013 2014 2015 2013 2014 2015 Inverse Probability Weighting
0.428***
(0.0341)
0.363*** (0.0343)
0.244*** (0.0421)
0.596*** (0.0519)
0.519*** (0.0597)

0.368*** (0.0493) Double-Robust 0.409***
(0.0355)
0.324*** (0.0367)
0.222*** (0.0446)
0.531*** (0.0588)
0.479*** (0.0649)

0.335*** (0.0455) Note: Standard errors in parentheses. *$p < 0.05$. **$p < 0.01$. ***$p < 0.001$.

Table 4. ATET estimation new bicentennial school.

ATET - Language eighth grade ATET - Math eighth grade Method 2013 2014 2015 2013 2014 2015
Inverse probability weighting 0.446***
(0.0683)
0.422*** (0.0587)
0.347*** (0.0447)
0.642*** (0.0832)
0.647*** (0.0814)

0.479*** (0.0687) Double-Robust 0.430***
(0.0658)
0.382*** (0.0591)
0.319*** (0.0455)
0.571** (0.0867)
0.603*** (0.0788)

0.443*** (0.0668) Note: Standard errors in parentheses. *$p < 0.05$. **$p < 0.01$. ***$p < 0.001$.

the lower results in 2015 may have been influenced by the loss of support from the Min- istry of Education starting in 2014 after a new government administration eliminated the pedagogical unit in charge of supporting Bicentennial schools. This meant that in 2014, schools did not administer the diagnostic test or the follow-up tests to check if learning was being achieved, and they also did not receive pedagogical advice. Nonetheless, it should be noted that the effects identified in 2015, while lower than those of previous years, were still positive and, at least in the case of mathematics, large.

Finally, all estimations showed that the Bicentennial schools were relatively more effective in boosting math scores than language scores. We believe that the nature of teaching and learning those subjects may explain the differences. For instance, language may be harder to teach, and literacy development may be influenced to a greater extent by the home literacy environment than math. Concretely, the Bicentennial math program offered a clearer strategy and more consistent teaching methods by rearranging the con- tents of the seventh- and eighth-grade curricula. Similar improvements were not offered for language.

Tables 4 and 5 show estimates by type of Bicentennial school: new or restructured, respectively. The two methods support the fact that new Bicentennial schools obtain better results than restructured schools. Differences are clearer for the second and third cohorts. Nonetheless, even most estimations of ATET for restructured Bicentennial schools can be considered large.

Robustness checks

Table 6 shows the results of the falsification test. When studying the SIMCE fourth-grade test-score "placebo" outcome, the estimated treatment effect is always non-significant and close to zero in absolute value, across methods and cohorts. Although other factors can still influence the treatment and the potential outcomes, the falsification test provides good evidence that the unconfoundedness assumption is likely to hold.

Figures 1 and 2 display CIC estimates. We found that the CIC results were consistent with previous estimates: Average quantile effects ranged from 0.27 to 0.44 *SD* for

Table 5. ATET estimation restructured bicentennial school.

ATET - Language eighth grade ATET - Math eighth grade Method 2013 2014 2015 2013 2014 2015 Inverse probability weighting
0.420***
(0.0414)
0.334*** (0.0386)
0.194** (0.0611)
0.573*** (0.0759)
0.453*** (0.0799)
0.314*** (0.0554) Double-Robust 0.398***
(0.0472)
0.294*** (0.0391)
0.176** (0.0551)
0.509*** (0.0765)
0.414*** (0.0843)
0.283*** (0.0673) Note: Standard errors in parentheses. *$p < 0.05$. **$p < 0.01$. ***$p < 0.001$.

Table 6. Falsification test.

ATET - Language fourth grade ATET - Math fourth grade Method 2013 2014 2015 2013 2014 2015
Inverse probability weighting −0.00160
(0.00617)
−0.00264
(0.00488)
0.00163 (0.00432)
−0.00348
(0.00561)
−0.000882 (0.00582)
0.00333 (0.00497) Double-Robust −0.00427 (0.0108)
−0.00783
(0.00955)
0.00385 (0.00649)
0.00377 (0.00790)
0.00631 (0.00816)
0.00359 (0.00606) Note: Standard errors in parentheses. *$p < 0.05$. **$p < 0.01$. ***$p < 0.001$.

language, and from 0.32 to 0.53 *SD* for math. Unlike propensity score methods, CIC results were similar for the 2013 and the 2014 cohorts, but displayed an important decline in the 2015 cohort. It is interesting to note that the program has a greater effect on the lower quantiles, particularly for language test scores. For instance, for quantile Q1 we find effects of around 1 *SD* in two cohorts, implying effects 4 times greater than for quantile Q9. In contrast, math test scores do not change significantly across quantiles.

Since the CIC method takes into account the distribution of unobservables (under some assumptions) and controls for changes in peer composition in terms of parents' education, expectations, and income, it is allowing us to control, to some extent, for the share of the peer effects that may be responsible for the better results at Bicentennial schools. Finally, we compared the performance of Bicentennial schools to Emblematic schools. Emble- matic students applied to their schools with fifth- and sixth-grade GPAs virtually identical to those of Bicentennial students, but they obtained, on average, test scores in fourth grade 0.5 *SD* higher than Bicentennial students. Moreover, family income and parents' education were higher in Emblematic schools. Therefore, this control group helped us control for unobservables, such as

student and parent motivation, because the edu- cational projects are similar: academic oriented with a focus on high achievement to earn a place at the best universities.

Figure 1. CIC estimates language.

Figure 2. CIC estimates math.

Table 7. ATET estimation bicentennial school (control group: Emblematic schools).

ATET - Language eighth grade ATET - Math eighth grade Method 2013 2014 2015 2013 2014 2015 Inverse Probability Weighting
0.352 **
(0.128)
0.310*** (0.0915)
0.282** (0.0877)
0.377 ** (0.134)
0.520** (0.186)

0.301* (0.121) Double-Robust 0.405***

(0.0998)
0.290*** (0.0713)
0.319** (0.0998)
0.447*** (0.101)
0.500*** (0.127)

0.332** (0.112) Note: Standard errors in parentheses. *$p < 0.05$. **$p < 0.01$. ***$p < 0.001$.

Table 7 shows that Bicentennial schools perform better than Emblematic schools. For language, estimates fluctuate between 0.4 and 0.3 *SD*; for math, they range from 0.5 to 0.3 *SD*. In fact, these latter results are similar to those obtained with the original control group. Certainly, one concern about using Emblematic schools as a control group is that some of these students missed school days because they joined the 2011–2013 student protests about the quality of education. Obviously, protests can negatively affect performance in the standardized tests, but it is difficult to argue that they systematically affected perform- ance three years in a row, especially considering that there were no long-lasting student protests after 2013.

Discussion

We have estimated the ATET effect of attending schools that participated in the Bicenten- nial schools program. Propensity score estimates ranged from 0.33 to 0.6 *SD* in math, and from 0.22 to 0.43 *SD* in language, showing great benefits for students who were admitted to these schools. We confirmed that our results are consistent through additional robust- ness checks. In addition, our results show a downward trend, namely, 2013 results are higher than 2015 results both in language and in math. Some factors could have

influenced these results, but we suggest that the loss of pedagogical support from the Ministry of Education could have affected the performance of Bicentennial schools, mainly in the 2015 cohort. This should be investigated further through future research.

Although we are not identifying each of the mechanisms explaining the results of the Bicentennial program, descriptions of the implementation of the program (Carrasco et al., 2014) may shed light on some. First, the educational project is academic oriented and explicitly states that one objective of the program is to prepare students to undertake higher education. This would naturally affect performance in a standardized test such as SIMCE. As a matter of fact, the program offered, through 2014, a systematic assessment of student learning that has been evaluated positively by school principals and teachers. Bicentennial schools also provide support to students who struggle with some subjects, in the form of extra classes and tutorials. In fact, the lowest achievers (among the students attending the Bicentennial schools) are the ones who reap the greatest rewards from attending a Bicentennial school. There is also more evidence corroborating that the schools achieved desirable results over and above standardized test scores. For instance, the attendance level

improved and the attrition rate decreased 5 percentage points once the program started. This may be a demonstration of a greater family commitment; in fact, some families expressed satisfaction with the new school regime even though the schools were currently more demanding.

Another mechanism that might be explaining these results is the peer effect. Duflo, Dupas, and Kremer (2011) reinforce the idea that tracking by ability is not bad per se since students who belong to a "less disperse" class achieve better results than students who are randomly assigned to a class. Nonetheless, since we find that Bicentennial schools perform better than the at least equally selective Emblematic schools, it would be difficult to argue that the academic results of Bicentennial schools are driven by aca- demic selection and peer effects only.

The large effect of attending a Bicentennial school is comparable to the results found for a privately run network of schools in Chile, as described in Henríquez et al. (2012). Most of the schools in the network are free of charge and do not select explicitly, but there may be some self-selection at play. Resorting to the qualitative assessment that was made of these schools, it is possible to find common areas between this program and the Bicenten- nial program. First, both are greatly focused on student learning and they seem to share similar institutional goals. Second, the academic standards are clearly defined in both. Third, student performance is monitored by a centralized team that identifies the critical points in the student learning process and finds solutions in conjunction with the schools' principals.

Implications for policy, practice, and future research

This study complements previous studies and confirms the productivity differences between restructured and new Bicentennial schools found in Alcalde (2014). One plausible reason for the relative advantage of new Bicentennial schools is team selection. Loeb, Kalogrides, and Béteille (2011) found that effective schools attract and hire more effective teachers when vacancies arise. This may be the case for new Bicentennial schools, since they could hire an entirely new staff in a context where profiles and job con- ditions and expectations were exceptionally transparent. Some schools, for example, were

able to stress the academic achievement orientation of the program by offering contracts where a fraction of the salary was contingent to the students' academic results. Further research is desirable to further our understanding of these mechanisms.

From a policy perspective, the program seems to add value to a downtrodden Chilean public education system (Carrasco et al., 2014), and to give more chances for public school students all over the country (not only the capital) to attend selective universities (Rivera & Guevara, 2017). Even though the program could not employ a full-blown student selection (as demand was often lower than spots available), students at least self-selected into the Bicentennials, making the project selective. The positive outcomes obtained, even after accounting for these selection issues, indicate that selective secondary schools in develop- ing countries may be more effective than in developed countries because of the lack of quality alternatives.

It must be noted, nonetheless, that our estimates provide only a short-term effect of the program, so school performance should be tracked to evaluate whether the program's strategies are sustainable over time and in other settings.

Notes

1. A *comuna* is the smallest administrative division in the country, and its local authority is the mayor. The mayor is responsible for public education. A "region" is the largest administrative division. 2. Unlike Emblematic schools located in

Santiago, the nation's capital, the program was widespread in the country. 3. The main source of missing data arises from incomplete parents' questionnaires. To check whether this is an issue, we compared, separately for treatment and control groups, the averages of several variables with and without missing data: GPA in fifth and sixth grades and math and language SIMCE test scores in fourth grade. We observed that missing obser- vations for the control group display, on average, GPA and SIMCE scores that are equal or lower than the control group used for the analyses. This would suggest that our estimates could represent a lower bound to the real impact of the program. On the other hand, we did not observe significant differences when performing the tests for the treatment group. Detailed estimates are available upon request. 4. SIMCE scores are divided into three categories: Insufficient, Elementary, and Appropriate. 5. US$ 1 = 500 Chilean pesos in 2010. 6. We implemented the propensity score using the PSCORE command (Becker & Ichino, 2002). The advantage of using PSCORE is the balanced covariate obtained across treatment and non-treatment groups and within strata, and the resulting optimal balance (Garrido et al., 2014).

Acknowledgements

We thank Mauricio Bravo and Juan Pablo Couyoumdjian for their participation in the UDD Bicente- nario interfaculty research project, and their comments and revisions of several drafts of this work. We also thank the Chilean Ministry of Education for facilitating us with the SIMCE data. This research is the sole responsibility of the research team and should not be interpreted as representative of the Ministry of Education's view.

Disclosure statement

No potential conflict of interest was reported by the authors.

Notes on contributors

Pablo Araya is currently advisor of the Undersecretary of Education of Chile. He was previously Direc- tor of Education at the Municipality of Melipilla, Chile. The focus of his research is the impact evalu- ation and the districts effectiveness in education. He holds a Master in Public Administration in Economics and Public Policy from The London School of Economics and Political Science. *Francisca Dussaillant* is currently president of the Chilean Society for Public Policy. She was pre- viously Head of Social Policy at the Chilean Ministry of Finance, Director of the Centrer for Public Policy at Universidad del Desarrollo, Counselor at the Chilean National Education Council, and UNDP consultant. She holds a PhD in Economics from Universidad Católica de Chile and an MA in Education from University of North Carolina, at Chapel-Hill.

ORCID

Francisca Dussaillant http://orcid.org/0000-0002-9687-4508

References

Abdulkadiroğlu, A., Angrist, J., & Pathak, P. (2014). The elite illusion: Achievement effects at Boston and New York exam schools. *Econometrica*, *82*(1), 137–196. doi:10.3982/ECTA10266 Adelson, J. L. (2013). Educational research with real-world data: Reducing selection bias with propen- sity score analysis. *Practical Assessment, Research & Evaluation*, *18*(15). Retrieved from http:// pareonline.net/getvn.asp?v=18&n=15 Alcalde, P. (2014). *Selective schools and the effect of class composition: Evaluation of the Bicentenario Schools in Chile*. Retrieved from https://editorialexpress.com/cgi-bin/conference/download.cgi? db_name=SECHI2014&paper_id=23 Allende, C. (2015). *Efectividad de los liceos públicos de excelencia en Chile* [Emblematic schools' effec- tiveness in Chile] (Master's Thesis,

Universidad de Chile). Retrieved from http://repositorio.uchile. cl/bitstream/handle/2250/134591/Efectividad de los liceos públicos de excelencia en Chile.pdf? sequence=1&isAllowed=y Athey, S., & Imbens, G. W. (2006). Identification and inference in nonlinear difference-in-differences

models. *Econometrica*, *74*(2), 431–497. doi:10.1111/j.1468-0262.2006.00668.x Becker, S. O., & Ichino, A. (2002). Estimation of average treatment effects based on propensity scores.

The Stata Journal, *2*(4), 358–377. doi:10.1177/1536867X0200200403 Bravo, M., Lavín, J., Ruiz de Viñaspre, J., & Wilkins, A. (2016). *Liceos de Excelencia Bicentenario: Vía rápida de inclusión y movilidad social* [Bientenario schools for excellence: Rapid path for inclusion and social mobility]. Santiago: RIL Editores. Bucarey, A., Jorquera, M., Muñoz, P., & Urzúa, S (2014). El efecto del Instituto Nacional. Evidencia a partir de un diseño de regresión discontinua [The effect of attending an elite public high- school in Chile. The case of the Instituto Nacional]. *Estudios Públicos*, *133*, 37–68. Retrieved from https://www.cepchile.cl/el-efecto-del-instituto-nacional-evidencia-a-partir-de-un-diseno-de/cep/ 2016-03-04/100601.html Carrasco, A., Gutierrez, G., Bogolasky, F., Rivero, R., & Zarhi, M. (2014). *Análisis del estado de implementación del programa "Liceos Bicentenario de Excelencia"* [Analysis of the implementation status of the "Bicentennial Schools for Excellence" program]. Retrieved from http://www.dipres. gob.cl/597/articles-139743_informe_final.pdf Cerulli, G. (2015). *Econometric evaluation of socio-economic programs: Theory and applications*.

Heidelberg: Springer. doi:10.1007/978-3-662-46405-2 Clark, D. (2010). Selective schools and academic achievement. *The B.E. Journal of Economic Analysis &*

Policy, *10*(1). doi:10.2202/1935-1682.1917 Clark, D., & Del Bono, E. (2016). The long-run effects of attending an elite school: Evidence from the United Kingdom. *American Economic Journal: Applied Economics*, *8*(1), 150–176. doi:10.1257/app. 20130505

Dobbie, W., & Fryer, R. G., Jr. (2014). The impact of attending a school with high-achieving peers: Evidence from the New York City exam schools. *American Economic Journal: Applied Economics*, *6*(3), 58–75. doi:10.1257/app.6.3.58 Duflo, E., Dupas, P., & Kremer, M. (2011). Peer effects, teacher incentives, and the impact of tracking: Evidence from a randomized evaluation in Kenya. *American Economic Review*, *101*(5), 1739–1774. doi:10.1257/aer.101.5.1739 Dustan, A. (2010). *Have elite schools earned their reputation? High school quality and student tracking in*

Mexico City. Retrieved from https://are.berkeley.edu/fields/ae/dustan.pdf Eyzaguirre, S. (2016). *¿Por qué liceos de excelencia?* [Why excellence-selective schools?]. Retrieved from

https://www.cepchile.cl/cep/site/artic/20161219/asocfile/20161219124144/pder448_seyzaguirre. pdf Fontaine, A., & Urzúa, S. (2014). *Selección, diversidad y libertad de enseñanza* [Selection, diversity and freedom of education] (Working Paper No. 10). Retrieved from Centro Latinoamericano de Políticas Económicas y Sociales – CLAPES website: http://www.clapesuc.cl/assets/uploads/2016/ 02/doct-10-vfinal-seleccion-diversidad-y-lib-ensenanza_v13_final_clapes_final-su-y-af.pdf Galindo-Rueda, F., & Vignoles, A. (2005). *The heterogeneous effect of selection in secondary schools: Understanding the changing role of ability*. Retrieved from Centre for the Economics of Education, London School of Economics and Political Science website: http://eprints.lse.ac.uk/ 19440/1/The_Heterogeneous_Effect_of_Selection_in_Secondary_Schools_Understanding_the_ Changing_Role_of_Ability.pdf Garrido, M. M., Kelley, A. S., Paris, J., Roza, K., Meier, D. E., Morrison, R. S., & Aldridge, M. D. (2014). Methods for constructing and assessing propensity scores. *Health Services Research*, *49*(5), 1701–1720. doi:10.1111/1475-6773.12182 González, F., & Johnson, E. (2018). Políticas de inclusión universitaria y comportamiento estratégico en educación secundaria [University inclusion policies and strategic behaviour in secondary edu- cation]. *Estudios Públicos*, *149*, 41–73. https://www.cepchile.cl/cep/site/docs/20180502/ 20180502104745/rev149_fgonzalezyejohnson.pdf Heinrich, C., Maffioli, A., & Vázquez, G. (2010). *A primer for applying propensity-score matching*.

Retrieved from http://publications.iadb.org/handle/11319/1681 Henríquez, F., Lara, B., Mizala, A., & Repetto, A. (2012). Effective schools do exist: Low-income chil- dren's academic performance in Chile. *Applied Economics Letters*, *19*(5), 445–451. doi:10.1080/ 13504851.2011.583208 Hsieh, C.-T., & Urquiola, M. (2006). The effects of generalized school choice on achievement and stra- tification: Evidence from Chile's voucher program. *Journal of Public Economics*, *90*(8–9),

1477–1503. doi:10.1016/j.jpubeco.2005.11.002 Imbens, G. W., & Wooldridge, J. M. (2009). Recent developments in the econometrics of program

evaluation. *Journal of Economic Literature*, *47*(1), 5–86. doi:10.1257/jel.47.1.5 Jackson, C. K. (2010). Do students benefit from attending better schools? Evidence from rule-based student assignments in Trinidad and Tobago. *The Economic Journal*, *120*(549), 1399–1429. doi:10. 1111/j.1468-0297.2010.02371.x Jackson, C. K. (2013). Can higher-achieving peers explain the benefits to attending selective schools? Evidence from Trinidad and Tobago. *Journal of Public Economics*, *108*, 63–77. doi:10.1016/j. jpubeco.2013.09.007 Khandker, S. R., Koolwal, G. B., & Samad, H. A. (2009). *Handbook on impact evaluation: Quantitative*

methods and practices. Washington, DC: The World Bank. doi:10.1596/978-0-8213-8028-4 Koch, A., Nafziger, J., & Nielsen, H. S. (2015). Behavioral economics of education. *Journal of Economic*

Behavior & Organization, *115*, 3–17. doi:10.1016/j.jebo.2014.09.005 Koljatic, M., & Silva, M. (2010). Algunas reflexiones a siete años de la implementación de la PSU [Some reflections 7 years after the implementation of the PSU – University entrance test]. *Estudios Públicos*, *120*. Retrieved from https://www.cepchile.cl/cep/estudios-publicos/n-111-a-la-120-2008-2011/algunas-reflexiones-a-siete-anos-de-la-implementacion-de-la-psu Lara, B., Mizala, A., & Repetto, A. (2011). The effectiveness of private voucher education: Evidence from structural school switches. *Educational Evaluation and Policy Analysis*, *33*(2), 119–137. doi:10.3102/0162373711402990

Loeb, S., Kalogrides, D., & Béteille, T. (2011). *Effective schools: Teacher hiring, assignment, development, and retention* (NBER Working Paper No. 17177). Cambridge, MA: National Bureau of Economic Research. doi:10.3386/w17177 Manning, A., & Pischke, J.-S. (2006). *Comprehensive versus selective schooling in England and Wales: What do we know?* Retrieved from Centre for the Economics of Education, London School of Economics and Political Science website: http://eprints.lse.ac.uk/19417/1/Comprehensive_Versus_Selective_Schooling_in_England_and_Wales_What_Do_We_Know.pdf Manríquez, B. (2016). *Efectos de los liceos públicos emblemáticos sobre el desempeño escolar el caso del Liceo N° 1 de Niñas Javiera Carrera* [Emblematic schools' effects: The case of the only-girls High School N°1] (Master's Thesis, Universidad de Chile). Retrieved from http://repositorio.uchile.cl/ handle/2250/138585 Melly, B., & Santangelo, G. (2015). *The changes-in-changes model with covariates*. Retrieved from

http://www.hec.unil.ch/documents/seminars/deep/1591.pdf Murnane, R. J., & Willett, J. B. (2010). *Methods matter: Improving causal inference in educational and*

social science research. New York, NY: Oxford University Press. Organisation for Economic Co-operation and Development. (2013). *PISA 2012 results: What makes schools successful? Resources, policies and practices (Volume IV)*. Retrieved from https://www. oecd.org/pisa/keyfindings/pisa-2012-results-volume-IV.pdf Organisation for Economic Co-operation and Development. (2014). *PISA 2012 results in focus: What 15-year-olds know and what they can do with what they know*. Retrieved from http://www.oecd. org/pisa/keyfindings/pisa-2012-results-overview.pdf Pop-Eleches, C., & Urquiola, M. (2013). Going to a better school: Effects and behavioral responses.

American Economic Review, *103*(4), 1289–1324. doi:10.1257/aer.103.4.1289 Rivera, M., & Guevara, J. (2017). *Selección en la educación secundaria y acceso a universidades de elite por tipos de establecimiento en Chile: Análisis de la cohorte 2016* [Secondary schools' selection and access to selective universities by type of schools: 2016 cohort analysis] (Working Paper No.8). Santiago: Centro de Estudios Mineduc, Ministerio de Educación. Retrieved from https:// centroestudios.mineduc.cl/wp-content/uploads/sites/100/2017/06/DctoTrabajo8-MRivera.pdf Schiefelbein, E., Leiva, R., & Schiefelbein, P. (2004). Closing the math learning gap between Chile and developed countries. *Journal of International Cooperation in Education*, *7*(1), 99–114. Retrieved from http://home.hiroshima-u.ac.jp/cice/wp-content/uploads/2014/03/7-1-10.pdf United Nations Educational, Scientific and Cultural Organization. (2015). *Informe de resultados TERCE: Logros de aprendizaje* [TERCE results report: Learning achievements]. Retrieved from http:// unesdoc.unesco.org/images/0024/002435/243532S.pdf