This article was downloaded by: [Universidad de Chile]
On: 03 February 2014, At: 11:54
Publisher: Routledge
Informa Ltd Registered in England and Wales Registered Number: 1072954 Registered office: Mortimer House, 37-41 Mortimer Street, London W1T 3JH, UK

The Journal of Educational Research
Publication details, including instructions for authors and subscription information:
http://www.tandfonline.com/loi/vjer20

Supporting Instructional Improvement in Low-Performing Schools to Increase Students' Academic Achievement

Cristián Bellei [a]

[a] University of Chile , Chile

Published online: 12 Feb 2013.

To cite this article: Cristián Bellei (2013) Supporting Instructional Improvement in Low-Performing Schools to Increase Students' Academic Achievement, The Journal of Educational Research, 106:3, 235-248, DOI: 10.1080/00220671.2012.687788

To link to this article: http://dx.doi.org/10.1080/00220671.2012.687788

PLEASE SCROLL DOWN FOR ARTICLE

Taylor & Francis makes every effort to ensure the accuracy of all the information (the "Content") contained in the publications on our platform. However, Taylor & Francis, our agents, and our licensors make no representations or warranties whatsoever as to the accuracy, completeness, or suitability for any purpose of the Content. Any opinions and views expressed in this publication are the opinions and views of the authors, and are not the views of or endorsed by Taylor & Francis. The accuracy of the Content should not be relied upon and should be independently verified with primary sources of information. Taylor and Francis shall not be liable for any losses, actions, claims, proceedings, demands, costs, expenses, damages, and other liabilities whatsoever or howsoever caused arising directly or indirectly in connection with, in relation to or arising out of the use of the Content.

This article may be used for research, teaching, and private study purposes. Any substantial or systematic reproduction, redistribution, reselling, loan, sub-licensing, systematic supply, or distribution in any form to anyone is expressly forbidden. Terms & Conditions of access and use can be found at http://www.tandfonline.com/page/terms-and-conditions

The Journal of Educational Research, 106:235–248, 2013
Copyright © Taylor & Francis Group, LLC
ISSN: 0022-0671 print / 1940-0675 online
DOI:10.1080/00220671.2012.687788

Routledge
Taylor & Francis Group

Supporting Instructional Improvement in Low-Performing Schools to Increase Students' Academic Achievement

CRISTIÁN BELLEI
University of Chile, Chile

Downloaded by [Universidad de Chile] at 11:54 03 February 2014

ABSTRACT. This is an impact evaluation of the Technical Support to Failing Schools Program, a Chilean compensatory program that provided 4-year in-school technical assistance to low-performing schools to improve students' academic achievement. The author implemented a quasi-experimental design by using difference-in-differences estimation combined with propensity scores matching procedures to estimate treatment effects. The main findings were the following: (a) the program had positive effects on fourth-grade students' achievement in both language and mathematics; (b) program effect size was 0.23 standard deviations, and not sensitive to control for covariates; (c) there were larger effects for students in the middle part of the students' test-score distribution; (d) after the intervention had ceased, the program impact declined rapidly; and (e) the program reduced grade retention by 1.5 percentage points.

Keywords: at-risk students, compensatory education, professional development, program evaluation, urban education

The effectiveness of alternative school-improvement proposals, particularly compensatory programs that are highly focused on student academic achievement, is the subject of much debate. Compensatory programs are regarded as alternatives to more controversial educational policies, such as modifying the nature of school governance or the student body composition of a school, and they have been a key component of the educational-equity agenda in many parts of the world, including the United States and Latin America. In this article I report an impact evaluation of the Technical Support to Failing Schools (TSFS) Program, a recent Chilean compensatory program that provided 4-year in-school technical assistance to low-performing schools to improve students' academic achievement.

Characteristically, most compensatory programs in education are temporally limited and targeted at specific student populations that possess a common disadvantage considered to hinder their educational success (Bouveau, 2004). Some of the most well-known compensatory programs in education include Educational Priority Areas in England, Better Schooling for Educationally Deprived Students (Title I) in the United States, and Priority Education Zones in France. While some compensatory programs simply supply additional financial resources to their target populations, others provide more complex interventions, including supplementary teacher training, enhanced pedagogic materials, and extra school services, among others (Gajardo, 2004; Winkler, 2000).

The impact of compensatory educational programs on student academic achievement has proven very difficult to evaluate in an unbiased fashion. The main challenge is providing a compelling counterfactual—that is, what the distribution of student achievement would have been in the absence of the intervention. Often, the target population is unique or fully covered by the program, eliminating the possibility of obtaining a reasonable control group (Rossi, Lipsey, & Freeman, 2003; Weiss, 1998).

The most relevant compensatory program in the United States has been Title I of the Elementary and Secondary Education Act (Borman & D'Agnostino, 2001; Kosters & Mast, 2003).[1] Although, as previous mentioned, several European countries have also implemented compensatory educational programs, these are typically interventions that are more structured and that have been implemented in educational systems in which state authorities have more direct control over the schools and the program implementation. In contrast, Chile possesses a highly decentralized educational system and the TSFS Program (the focus of this study) was implemented simultaneously by several autonomous nongovernmental institutions. In this respect, the evaluation of Title I offers a more appropriate point of comparison for an impact evaluation of the TSFS Program.

Many studies have attempted to evaluate the impact of Title I programs on subsequent student achievement. Unfortunately, most of them have suffered from serious methodological weaknesses: Many did not identify—or even include—an

Address correspondence to Cristián Bellei, Center for Advanced Research in Education, University of Chile, Periodista Jose Carrasco Tapia, 75, Santiago, 8330014 Chile. (E-mail: cbellei@uchile.cl)

Downloaded by [Universidad de Chile] at 11:54 03 February 2014

adequate control group, did not use comparable outcome measures, and most were not conducted in nationally representative samples (Borman & D'Agostino, 2001; Kosters & Mast, 2003; McDill & Natriello, 1998). These limitations have induced great heterogeneity in the estimated sizes of the obtained program effects. For example, Kosters and Mast (2003) reported 80 estimates of Title I impact on students' National Assessment of Educational Progress scores, with an effect size ranging from −0.6 to +0.3 of a standard deviation.

The Sustaining Effects Study (Carter, 1984) and the Prospects Study (Borman, D'Agostino, Wong, & Hedges, 1998; Puma et al., 1997) are two of the most relevant studies of the impact of Title I. Both are longitudinal studies of nationally representative samples of roughly 120,000 and 40,000 primary students, respectively. The Sustaining Effects Study found a statistically significant effect of the Title I intervention on students' mathematics achievement in Grades 1–6, and on their language achievement in Grades 1–3; it also found some evidence of greater effectiveness of the intervention among moderately disadvantaged students (Carter, 1984). Additionally, the Prospects Study found that Title I funded interventions were not sufficient to close the achievement gap between Title I and nondisadvantaged students (Puma, 1999; Puma et al., 1997); however, when compared to similar nonparticipating students, Title I students (especially those with the greatest advantages) showed greater achievement gains (Borman, et al., 1998). Complementarily, one meta-analysis of methodologically rigorous Title I evaluations that were completed between 1966 and 1999 estimated an average effect size of slightly more than one tenth of a standard deviation in student achievement—mainly in language and mathematics in primary school (Borman & D'Agostino, 2001). These authors also reported that the estimated effect size was greater in mathematics and for children in the lower grades. Finally, an additional meta-analysis of 232 impact evaluations of 29 comprehensive school reforms, also funded by Title I, reported an average effect size of about 0.15 of a standard deviation, and the mean estimated effect size among the programs having the strongest evidence ranged from 0.15 to 0.21 of a standard deviation (Borman, Hewes, Overman, & Brown, 2003).

Most compensatory programs implemented in Latin American countries have not been oriented directly towards the improvement of students' academic achievement (Gajardo, 2004). Accordingly, prior research on compensatory programs in Latin America has been mainly focused on increasing educational coverage in the population of eligible children and improving the conditions of the schools they attend, instead of focusing directly on the improvement of student's academic performance. Within these constraints, limited evidence points to the very low impact, if any, of compensatory programs on student academic achievement (Reimers, 2000a, 2000b; Winkler, 2000).

In Chile, there have been few methodologically rigorous evaluations of the impact of compensatory programs on students' academic achievement. Chay, McEwan, and Urquiola (2005) used regression-discontinuity methods to estimate the causal impact of the P-900 program[2] on Grade 4 mathematics and Spanish language achievement between 1990 and 1992 and found that it was considerable, with an effect size of about one fifth of a standard deviation. A second more comprehensive evaluation of the impact of the same program on Grade 4 average mathematics and Spanish language achievement used a difference-in-differences design with multiple matching comparison groups. Depending on the particular program cohort, evaluated between 1990 and 1999, the estimated effect size ranged from −0.09 to 0.36 of a standard deviation (Asesorías para el Desarrollo & Santiago Consultores, 2000).

Overall, previous studies have found generally that compensatory educational programs have had a positive, but small to moderate, impact on students' academic achievement. The effects have been more consistent on mathematics achievement than on language, and only apply to the early primary grades. There is also evidence that programs have been more effective for moderately low-performing students than for extremely low-performing students.

Recently, some governments have adopted new strategies to improve the effectiveness of compensatory programs (Bouveau, 2004; Reimers, 2000a). In particular, in 2002, the Chilean Government launched the TSFS Program, a compensatory program that used nongovernment educational experts to support failing schools in a highly decentralized educational system. Providing technical assistance to failing schools is also an educational intervention of burgeoning importance in U.S. education, within the context of the No Child Left Behind Act of 2001 (2002; Finnigan & O'Day, 2003; Laguarda, 2003; Sheekey, Cymrot, & Fauntleroy, 2005; Turnbull, White, Sinclair, Riley, & Pistorino, 2011). This practice was extensively disseminated among U.S. districts precisely after the 1994 reauthorization of Title I.[3]

In the present study I take advantage of the exogenous implementation of the TSFS compensatory program in Chile to obtain an unbiased estimate of its causal impact on students' academic outcomes.

Supporting Failing Schools in Chile: The TSFS Compensatory Program

Since the early 1980s, Chile has developed a nationwide market-oriented educational system and, starting in 1990, resources allocated to public education increased dramatically. Nevertheless, the level of educational inequality, as measured by the gap in average academic achievement between students of high and low socioeconomic status, has remained high (Bellei & Mena, 2000; Garcia-Huidobro & Bellei, 2003). As expected, Chilean policymakers now wonder why both market incentives and the provision of additional financial resources have been ineffective in improving the academic achievement of students in the most disadvantaged schools. One plausible hypothesis is that schools and teachers in these poor areas lack the

Downloaded by [Universidad de Chile] at 11:54 03 February 2014

professional capacities to react productively to market incentives and to take advantage of the additional resources (Bellei, 2003; Cox, 2003; Organization for Economic Cooperation and Development, 2004). The TSFS program was expected to tackle this issue by creating school-based programs of teacher professional development.

The goal of the TSFS Program was to improve the academic achievement of students in the failing elementary schools of Santiago by developing the professional capacities of teachers and the quality of school management. To this end, it provided external technical assistance to failing schools, mainly in the form of in-school teacher training.

The TSFS Program was based on four principles about school change. First, to improve student achievement in the target schools, teachers must alter their current teaching practices. Second, changing teaching practice is a school-level organizational challenge. Third, strong school leadership in addition to professional teamwork among teachers is required for planning and coordination. Finally, the quality of the social relationships among teachers and administrators in the target schools influence strongly the effectiveness of the school in improving instruction.

The TSFS Program was implemented in 70 publicly funded primary schools in the Santiago Metropolitan Area[4] starting at the beginning of the 2002 academic year. Schools remained in the program until the end of the 2005 academic year. The schools that were assigned to the program were those whose average scores on the Chilean National Testing System (SIMCE) assessments of Spanish language and mathematics, among fourth-grade students, was less than 230 at the end of the 1999 academic year.[5] Additional selection criteria (e.g., high grade retention rate, high dropout rate, medium or large school size, school socioeconomic status, and years of school participation in previous publicly funded school-improvement programs) were also applied.

The program was intended to bolster the achievement of children in the target schools up to fourth grade. Specifically, the objectives of the TSFS Program were to increase the language and mathematics achievement of students in targeted schools to the national average by the end of Grade 4, and to reduce the grade retention rate in participating schools from an average of 5% to less than 2% (Ministry of Education, 2006).

During the 4 years of program implementation, each participating school received support from external experts (drawn from universities and research centers) to improve teachers' classroom practices mainly from Grades 1–4, school organization/management, and school social climate (Sotomayor & Dupriez, 2007). Although there was some level of heterogeneity among consultant teams from different institutions, they all applied a similar program intervention. Basically, during TSFS implementation, external consultants provided target schools with the practical tools needed to trigger the school-improvement processes, starting with the design of a school-improvement plan. Particularly, external consultants provided detailed teaching materials, additional pedagogical training, in-school guidance to both teachers and principals, and help with the diagnosis and treatment of students with learning difficulties or behavioral problems (Ministry of Education, 2006).

Finally, a process evaluation conducted during the second year of implementation found that the program had been implemented satisfactorily in most of the participating schools and that some of the expected processes of change were identifiable, especially at the classroom level (Programa de las Naciones Unidas para el Desarrollo & Asesorías para el Desarrollo, 2004).

Method

In this paper I report on an impact evaluation of the TSFS Program. In order to make a causal inference about the relationship between program participation and the measured outcome, I used a difference-in-differences estimation strategy to investigate whether Chilean primary school students who attended targeted schools during the program implementation had higher academic achievement than comparable students who studied in the same schools prior to the implementation of the intervention (Meyer, 1995; Shadish, Cook, & Campbell, 2002; Weiss, 1998; Wooldridge, 2002). The character of the TSFS implementation allows me to design a quasi-experiment with which to evaluate program impact.

Implementation of the TSFS Program was restricted to the Santiago Metropolitan Area because of financial constraints and the government's wish to develop a pilot for similar intervention programs in other regions. Therefore, Santiago students attending program schools were arbitrarily exposed to program intervention, after the program was implemented, but not before. This arbitrariness provided exogenous variation in the primary-school experience among Chilean students and enabled me to estimate a first difference in academic achievement between children in the erstwhile treatment (postintervention) and control (preintervention) groups. Fortunately, the SIMCE databases also contain considerable information on the academic achievement of children in schools throughout other regions of the country. Consequently, by using a matching procedure based on preprogram information (including students' academic achievement and other school covariates), I was able to select a comparison group of observably equivalent primary schools in nonprogram metropolitan areas. Based on the test scores of students in these latter schools, I then estimated a second difference to represent any historical trend in student achievement that differentiated cohorts of students nationally, pre- and postimplementation. By subtracting the estimated second difference from the first, I could remove the effects of the secular trend from my estimate of the causal impact of program on student achievement.

Compared to previous research on compensatory programs, my study has several methodological strengths. First, in the context of a new policy that introduced an exogenous change among schools, I used propensity score matching to create a comparison group for purposes of estimating and removing secular trends in achievement over time. Second, the program was in place for a long period of time, affecting the evaluated cohort of students over 4 years, and therefore had the potential to lead to reasonably sized differences in student achievement. Third, I made comparisons by following the same group of primary schools over time (moreover, at the school level, I analyzed more than one outcome measure postprogram implementation). Fourth, outcome measures were available, at the student level, in both language and mathematics.

Data Sets

I analyzed data from the SIMCE-1999 and SIMCE-2005 databases (my pre and post measures, respectively).[6] SIMCE is the Chilean national testing system, and assesses primary school students in mathematics and language achievement at the end of Grade 4, every 3 years. In 1999 and 2005, 96% of Chilean fourth-grade students were assessed. Additionally, SIMCE data sets contain individual-level background information on the students provided by their parents for 85%–90% of the students assessed (depending on the particular covariate). The SIMCE-1999 and SIMCE-2005 data sets also contain information on selected characteristics of the schools that the students attend. Finally, supplementary school-level information (i.e., geographical location, enrollment, grade retention, and dropout rate) was obtained from the School Directory, an official yearly updated database. Thus, my study compares different cohorts of students within the same group of schools, analyzing student- and school-level data.

Sample

I implemented a matching procedure using propensity score methods to create a comparison group of schools in nonprogram metropolitan areas,[7] within which I could estimate my required second difference.[8] Specifically, I fitted a logistic regression model in order to estimate the schools' probabilities of being a member of the TSFS Program group for both program schools and for all potential comparison schools in the Valparaiso and Concepción areas. I estimated propensity scores for each school to predict the probability of program participation using predictors chosen applying the criteria explicitly stated by program authorities as covariates: the school average language/mathematics SIMCE-1999 score, the school grade retention rate (2001), the school dropout rate (2001), the school enrollment (2001), the school average socioeconomic status (1999), the school geographical location (urban/rural condition), the school type (public or private-subsidized condition) and the number of years that the school had participated in previous school improvement programs. The fitted probabilities estimated from this logistic regression model provide the school-level probabilities (propensities) of treatment.

Finally, each of the program schools[9] was matched pairwise with the nonprogram school having the most similar propensity score to create a comparison group of schools, the scores of whose students permitted me to estimate secular trends in achievement over the same time period as my first difference was estimated in the program schools. As a result, the program group was composed of 6,763 students who attended 69 primary schools that participated in the TSFS Program, and the comparison group comprised 6,341 students who attended 69 primary schools that did not participate in the TSFS Program.

Measures

In the Appendix I present formal definitions of all variables included in the present study. The variables involved in my difference-in-differences estimation strategy are defined as follows:

Outcome variables: (a) item response theory (IRT)–scaled mathematics SIMCE-2005 score and (b) IRT-scaled language SIMCE-2005 score. SIMCE scores summarize a student's degree of mastery of the National Curriculum Objectives, which are compulsory for all Chilean schools, and use IRT-equating to render scores vertically equitable from year to year.[10]

Question predictors: My principal question predictors are (a) PROGRAM, a dummy variable that indicated whether the student attended a program school (PROGRAM = 1) or a comparison school (PROGRAM = 0) and (b) POST, a dummy variable that indicated whether the student attended school after the program was introduced (POST = 1), or before program implementation (POST = 0).

Covariates: In order to test the sensitivity of my program-impact estimates, I also included in my analyses several student- and school-level covariates relevant to the prediction of student academic achievement (see the Appendix for a complete list of control variables used in the analyses).

Data Analysis

In my data analysis, I used a standard difference-in-differences strategy. The effect of the TSFS Program is contained within the difference in students' test scores between the SIMCE-2005 and SIMCE-1999 achievement scores, in the program schools (i.e., students who typically started Grade 1 in 2002 and 1996, respectively). However, this first difference may be contaminated by any secular trend that has differentially affected members of the two cohorts of students, not only in the program schools, but also in the rest of similar Chilean schools. To remove this historical component, I estimated a second difference between

Downloaded by [Universidad de Chile] at 11:54 03 February 2014

the SIMCE-2005 and SIMCE-1999 achievement scores in the comparison group. I implemented this difference-in-differences strategy in a multiple regression framework, as follows:

$$TestScore_{ij} = \beta_0 + \beta_1 PROGRAM_{ij} + \beta_2 POST_{ij} + \beta_3 PROGRAM_{ij}{}^{*}POST_{ij} + (\varepsilon_{ij} + \mu_j), \quad (1)$$

where $TestScore_{ij}$ is either the mathematics or language SIMCE-score of student i in school j, β_1 captures the net difference between program and comparison school students in the pre years, β_2 captures the net difference between students tested in 2005 and 1999 in comparison (nonprogram) schools, and the parameter associated with the two-way *PROGRAM*POST* interaction provides the difference-in-differences estimate of the impact of the program. Finally, ε_{ij} is an individual-level error term, and μ_j is the school-level error term. Thus, the proposed model accounts for the clustering of students at the school level, and provides an appropriate estimate of standard errors.[11] In Equation 1, if β_3 is positive and statistically significant ($p < .05$), I can conclude that the TSFS Program increased academic achievement among participating students.

Sensitivity Analysis

Introduction of covariates. Additional criteria (other than official criteria) and local circumstances may account for some part of the decision as to which particular schools entered into the TSFS Program. Thus, even after propensity score matching, program and comparison schools may still differ on additional covariates. Similarly, because the treatment condition was assigned at school level, students in the program and comparison schools, and pre- and posttreatment implementation students, may also differ on additional relevant covariates. Hence, in order to account for the possible remaining bias, I also fitted the following regressions:

$$TestScore_{ij} = \beta_0 + \beta_1 PROGRAM_{ij} + \beta_2 POST_{ij} + \beta_3 PROGRAM_{ij}{}^{*}POST_{ij} + \beta_4 X_j + (\varepsilon_{ij} + \mu_j); \quad (2)$$

$$TestScore_{ij} = \beta_0 + \beta_1 PROGRAM_{ij} + \beta_2 POST_{ij} + \beta_3 PROGRAM_{ij}{}^{*}POST_{ij} + \beta_4 Z_{ij} + (\varepsilon_{ij} + \mu_j); \quad (3)$$

$$TestScore_{ij} = \beta_0 + \beta_1 PROGRAM_{ij} + \beta_2 POST_{ij} + \beta_3 PROGRAM_{ij}{}^{*}POST_{ij} + \beta_4 X_j + \beta_5 Z_{ij} + (\varepsilon_{ij} + \mu_j); \quad (4)$$

where X_j is a vector of school covariates and Z_{ij} is a vector of student covariates. If the program-effect estimate is not impacted by the introduction of the covariates, I can conclude that the nonrandom assignment did not affect the original estimate.

Program impact on grade retention. An increase in grade retention in the program schools can also confound an estimate of program impact on student test scores, because low-performing students retained in Grades 1–3 are then not tested by SIMCE in Grade 4. In order to tackle this issue, I used the same difference-in-differences methodology to estimate the impact of the program on grade retention. Because the reduction of grade retention was also a TSFS Program objective, this estimate is of substantive interest as well.

Program effect at different levels of the achievement distribution. In evaluating compensatory programs (and more generally, school improvement programs), an additional objective is to estimate their impact on different student populations, particularly students with different academic skills. To determine whether the program effect differed at different levels of the distribution of students' academic achievement, I obtained difference-in-differences estimates of program effect at each decile of the students' test score distributions in both mathematics and language achievement.

Additional POST program measure. To explore the sustainability of the program effect, I also estimated program impact using Grade 4 SIMCE-2006 test scores as the POST program measure. Note that this involves a different cohort of students (who typically started Grade 1 in 2003), assessed 1 year after the program had finished.

Results

Validity of the Comparison Group

In the difference-in-differences design, the comparison group provides an empirical estimate of the expected difference in the outcome variables that the program group would have experienced in the absence of treatment—the so called secular trend. This estimate serves as the important second difference in the analysis. Thus, it is crucial to verify whether the matching procedure to identify the comparison schools worked with integrity when equating program and comparison groups. In Table 1, I present descriptive statistics on the program and comparison schools on selected variables measured at both the school and student level.

As expected, program and comparison schools are highly similar in most of the observed school-level characteristics (see the results of the t testing for mean differences, in the last column). Nevertheless, compared to the comparison schools, program schools included a slightly smaller proportion of public schools and aggregate SIMCE-1999 language test scores that were lower ($p < .1$). Additionally, both

Downloaded by [Universidad de Chile] at 11:54 03 February 2014

TABLE 1. Sample Statistics on Selected Variables, for Program and Comparison Schools

	Program			Comparison			
Variable	*n*	M	*SD*	*n*	M	*SD*	*t* (H_0: M difference = 0)
Outcome variables (school level)							
MATH_1999		214.26	12.09		213.33	13.37	−0.43
LANG_1999		208.23	12.55		212.52	15.34	1.80†
School-level covariates							
Language/Mathematics 1999 test score		211.25	11.64		212.93	13.13	0.80
Eligible		0.96	0.21		0.96	0.21	0.00
Urban		0.83	0.38		0.78	0.42	−0.64
Public		0.77	0.43		0.88	0.32	1.80†
SES_1999		4.59	0.86		4.54	1.04	−0.36
Enrollment_2001		462.99	243.75		416.36	273.51	−1.06
Grade Retention_2001		2.83	2.11		2.70	1.96	−0.39
Dropout_2001		2.96	4.04		2.16	3.64	−1.22
SES_2005		47.58	10.65		50.12	14.69	1.16
Student-level covariates							
Income		0.74	1.16		0.59	1.13	−7.16***
Retained		0.23	0.42		0.22	0.41	−2.11*
Father's education		8.30	3.29		8.31	3.33	0.18
Mother's education		8.01	3.18		8.16	3.18	2.56*
Schools	69			69			
Student minimum	5,670			5,427			
Student maximum	6,010			5,810			

$^{\dagger}p < .10$. $^{*}p < .05$. $^{***}p < .001$.

Downloaded by [Universidad de Chile] at 11:54 03 February 2014

groups of schools differ on three of the four student-level observed covariates (recall that the matching procedure was implemented at the school level): Students attending program schools tended to have higher levels of grade retention and less educated mothers ($p < .05$), but they also tended to have families with higher incomes ($p < .001$), than did comparison schools. To rule out the possibility that these small observed differences between groups produced bias in the results, I conducted two additional analyses. For a difference-in-differences estimate, the absolute observed differences between program and comparison groups are less relevant than the differences in changes between PRE and POST students within the program and comparison schools, which in this case involves two different cohorts of students. Therefore, if my assumption concerning the equality in expectation of students in the program schools pre- and postintervention is correct (correcting for any secular national trend estimated from the comparison schools over the same period), I would expect that there to be no difference on important covariates between these groups. I can test this assumption using the difference-in-differences strategy itself. Accordingly, in Table 2, I present difference-in-differences estimates obtained for four important student-level characteristics. Notice that the estimated difference-in-differences are close to zero over

TABLE 2. Sample Difference-in-Differences Estimates for Selected Student-Level Covariates

Variable	DD (POSTp-PREp) – (POSTc-PREc)	*t* (H_0: M difference = 0)	*n* schools	*n* students
Income	0.01	0.33	138	11,677
Retained	−0.02	−1.12	138	11,820
Father's education	0.12	0.91	138	11,097
Mother's education	0.28	2.36*	138	11,706

Note. Each coefficient was obtained from a different regression analysis.
$^{*}p < .05$.

Downloaded by [Universidad de Chile] at 11:54 03 February 2014

the period of analysis on all covariates except for mother's education, which while small, is positive and statistically significant ($p < .05$). This implies that—on average—the mother's education was increasing slightly faster in the program schools than in the comparison schools.

Finally, I conducted a test proposed by Winship and Morgan (1999) to verify whether small differences between both groups on unobserved variables can bias the program effect estimate. In this test, I fitted four regression models (baseline, with school covariates, with student covariates, and with school and student covariates) to data on only preprogram observations in both program and comparison schools (i.e., SIMCE-1999), to determine whether I could detect a false program effect before the introduction of the treatment. Because no treatment was actually applied at that period, the presence of a statistically significant pseudoprogram effect would suggest that initial differences between both program and comparison schools were impacting the outcome. I present the results of this test in Table 3. I detected no statistically significant differences in mathematics test scores in any of the four regression models. The findings for language scores were more complex. The first row of the table indicates that students in the comparison schools scored higher than students in program schools on language outcomes ($p < .05$). However, the results obtained in the fitting of Model 4, in which I added observed student- and school-level covariates (and also in Model 2, in which I included only school-level covariates), show that this difference is zero. In other words, in addition to the matching procedure, the inclusion of additional covariates succeeds in rendering an unbiased program impact estimate on language achievement.

On the basis of the previous analyses, I conclude that neither unobserved variables nor initial differences between program and comparison groups confounded my estimates of the causal effect of TSFS Program reported in the following sections.

TABLE 3. Average Differences in SIMCE-1999 Test Score Between Students Attending Program and Comparison Schools, Without, and with Selected School-Level and Student-Level Covariates

	Outcome variable	
	Mathematics	Language
Model 1: Without covariates	−0.17	−5.12*
Model 2: With school-level covariates	−0.07	−0.02
Model 3: With student-level covariates	−1.25	−6.06**
Model 4: With school- and student-level covariates	−0.65	−0.73
Maximum number of students	7,796	7,800

Note. Each coefficient was obtained from a different regression analysis. School-level covariates include Lang/Math 1999 test score, Eligible, Urban, Public, SES_1999, Enrollment_2001, Grade Retention_2001, Dropout_2001, MATH_1999, and LANG _1999. Student-level covariates include: Income, Retained, Father's education, and Mother's education.
*$p < .05$. **$p < .01$.

Difference-in-Differences Estimates of Program Effect

In Table 4, I present the results of fitting the four regression models in Equations 1–4 for both the mathematics and language outcomes. The estimated effect of program participation in both the mathematics and language, unadjusted for the impact of covariates (i.e., Model 1), is 10.57 and 11.77 SIMCE scores, respectively ($p < .001$). As part of the sensitivity analysis (also in Table 4), I introduced school-level covariates (Model 2) and student-level covariates (Model 3) for both the mathematics and language outcomes, and, finally, I controlled for school- and student-level covariates simultaneously (Model 4). School-level controls include all the variables used in the propensity score estimation, plus some relevant school characteristics measured in 2005 and the corresponding mathematics or language preprogram value. As shown in Table 4, the magnitude of the estimated program effect on both mathematics and language achievement remains unaffected and statistically significant ($p < .001$) in these three regression models.

These findings strongly support my conclusion that the TSFS Program has had a positive and statistically significant effect on students' academic achievement in mathematics and language academic achievement. After controlling for student and school characteristics, the estimated average program effect was about 10.84 in mathematics (ES = 0.23 standard deviations), and about 11.14 in language (ES = 0.23 standard deviations),[12] which represents a medium effect size.

Program Impact on Grade Retention

In Table 5, I present two difference-in-differences estimates of program impact on grade retention rate. Note that because grade retention rate is a school-level outcome, these estimates are based on school-level regression models: first, I estimated the causal impact of the program uncontrolled for other covariates (Model 1); then, I added several relevant school-level control variables (including school dropout which is a sensitive issue in this analysis; Model 2). In both models, program participation has a statistically significant negative effect on grade retention rate ($p < .05$). Note that the introduction of control variables increased the negative program effect estimates slightly.

My best estimate of the program impact, reported in Model 2, is −1.48 percentage points, which represents a

TABLE 4. Taxonomy of Fitted Multilevel Regression Models, Demonstrating the Relationship Between Student's Test Score (Mathematics and Language) and Program Participation, Controlling for Selected School-Level and Student-Level Covariates

	Mathematics				Language			
Variable	Model 1	Model 2	Model 3	Model 4	Model 1	Model 2	Model 3	Model 4
PROGRAM*POST	10.57***	10.21***	11.20***	10.84***	11.77***	11.46***	11.38***	11.14***
Program	−1.08	0.12	−2.04	−0.49	−6.24**	−1.15	−7.13***	−1.72
POST	−1.76	−1.80	−5.30***	−5.23***	8.89***	8.84***	5.48***	5.49***
Eligible		1.08		1.41		1.88		0.91
Lang/Math_1999		−0.21		−0.21		0.13		−0.00
Urban		−2.61		−1.83		−2.42		−0.74
Public		−1.66		−1.83		−2.05		−2.34
SES_1999		0.09		1.50†		0.39		1.73*
Enrollment_2001		−0.00		−0.00		−0.01		−0.01
Grade Retention_2001		1.19***		1.70***		1.13**		1.25**
Dropout_2001		−0.27		−0.35		−0.26		−0.34
Enrollment_2005		−0.00		−0.00		0.00		0.00
Grade Retention_2005		−0.27		−0.09		−0.23		0.04
Dropout_2005		−1.14***		−1.15***		−1.26***		−1.20***
SES_2005		−0.11		−0.06		−0.02		0.08
SES_MID		7.86*		2.64		7.28*		1.77
SES_MIDLOW		3.12		1.81		1.90		−0.76
PRE test Math or Lang		0.73***		0.63***		0.49***		0.49**
Income			1.29**	1.27**			1.08*	1.03*
Retained			−14.24***	−13.91***			−16.51***	−16.05***
Father's education			1.04***	1.02***			1.21***	1.20***
Mother's education			1.71***	1.69***			1.71***	1.68***
Intercept	216.53***	112.26***	199.77***	107.50***	216.12***	86.03***	199.16***	86.28***
R^2 between schools	1.7%	60.8%	21.6%	50.0%	0.9%	68.5%	27.1%	58.9%
R^2 within schools	0.5%	0.5%	5.1%	5.1%	2.7%	2.7%	8.2%	8.2%
n students	13,067	13,067	10,178	10,178	13,104	13,104	10,196	10,196

Note. Schools $n = 138$.
$^{\dagger}p < .10$. $^{*}p < .05$. $^{**}p < .01$. $^{***}p < .001$.

Downloaded by [Universidad de Chile] at 11:54 03 February 2014

medium to large effect size of 0.49 standard deviations in grade retention rate,[13] implying that the TSFS Program reduced grade retention among participating schools substantially. This is not only a positive program outcome, but is also proof that the estimated effect of the program on the academic achievement of students is not driven by a disproportionate increase in grade retention among participant schools.

Program Effect at Different Levels of the Achievement Distribution

To determine whether the program effect differed at different levels of the distribution of students' academic achievement, I obtained difference-in-differences estimates of the program effect at each decile of the students' test score distributions in both mathematics and language achievement, by implementing a quantile regression analysis. I only included school-level covariates in order to preserve the total sample (i.e., the analysis is equivalent to Model 2 in Table 4). I display the results in Figure 1 (see Table A2 in the Appendix for the exact coefficients). As shown, the estimated effect of the program was positive along the entire distribution of student achievement. In addition, the estimated effect size in mathematics and language achievement is somewhat larger in the middle portion of the students' test scores distribution (especially between deciles 2 and 7), ranging from 0.12 to 0.31 standard deviations in language, and from 0.12 to about 0.27 standard deviations in mathematics. This evidence suggests that the TSFS Program increased the average test scores of students in the participant schools by increasing the academic results for the general student population, and not by focusing only on one subgroup of students.

Sustainability of the Program Impact

Finally, a critical issue about effective compensatory educational programs is whether their positive results are preserved when the program intervention has ceased. To investigate this question, I fitted the four defined regression models using 1-year posttreatment data to estimate the program impact in a different cohort of students: fourth graders

Downloaded by [Universidad de Chile] at 11:54 03 February 2014

TABLE 5. Fitted School-Level Regression Models, Demonstrating the Relationship Between the Grade Retention Rate and Program Participation of Schools, Controlling for Selected School-Level Covariates

	Outcome: Grade Retention_2005	
Variable	Model 1	Model 2
Program impact	−1.28*	−1.48*
Intercept	3.73***	18.73***
Grade Retention_2001	0.72***	0.64***
Eligible		−2.06
Public		−1.49†
Urban		−0.67
SES_1999		0.20
SES_2005		0.00
Enrollment_2001		−0.00
Enrollment_2005		0.01
Dropout_2001		0.04
Dropout_2005		0.19
SES_MID		−0.67
SES_MIDLOW		−0.92
Lang/Math_1999		−0.02
Mathematics_2005		−0.02
Language _2005		−0.02
R^2	18.2%	32.9%
n schools	138	138

†$p < .10$. *$p < .05$. ***$p < .001$.

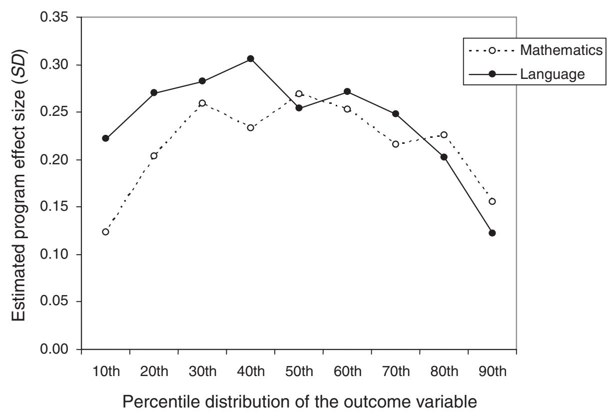

FIGURE 1. Estimated program effect at different levels of the students' academic achievement distribution. Difference-in-differences estimates of the program effect at the 10th, 20th, 30th, 40th, 50th, 60th, 70th, 80th, and 90th percentile of the students' test-score distributions in both mathematics and language. Outcome variable: SIMCE test scores 2005; student-level analysis (effect size calculated in standard deviation units). All coefficients are statistically significant ($p < .05$). School-level covariates include Lang/Math 1999 test score, Program, POST, Eligible, Urban, Public, SES_1999, Enrollment_2001, Grade Retention_2001, Dropout_2001, Enrollment_2005, Grade Retention_2005, Dropout_2005, SES_2005, SES_MID, SES_MIDLOW.

in 2006. I present the results in Table 6 for both mathematics (upper panel) and language (lower panel) test scores.

As shown, according to the results of all four regression models, the TSFS Program did not have an effect on students' language academic achievement 1 year after the conclusion of the program. Nevertheless, the estimates in Table 6 indicate consistently that the program did have a statistically significant positive effect on students' mathematics test scores. The estimated effect size of the program impact on mathematics achievement ranges from 0.08 to 0.09 standard deviations, depending on the model specification, which is a comparatively small effect size. Note that this program effect on mathematics results of the 2006 fourth-grade students is about one third of the estimated effect on the 2005 fourth-grade students, which I reported in Table 4. Taken as a whole, these findings suggest that the effect of the TSFS Program diminishes rapidly 1 year after the program has ended, and remains noticeable only for mathematics academic achievement.[14] In order to have a visual representation of the main research results, Figures 2 and 3 display the evolution of the estimated mean academic achievement (in mathematics and language, respectively) for students attending program and comparison schools. These graphs show a somewhat different pattern between both subjects: Whereas in mathematics the program group increased their academic achievement only during the TSFS Program intervention and then decreased it slightly, in language, program participants increased their academic achievement during the entire period but at a different rate. Therefore, the key to understand the difference in the estimated program impact between both subject matters in the follow-up assessment is the positive trend experienced by the comparison group in language test scores. Overall, Figures 2 and 3 demonstrate the high sensitivity of the students' academic achievement to the presence of the TSFS Program intervention.

Discussion

In this article I presented an impact evaluation of the Chilean TSFS Program, which provided 4-year external technical assistance (mainly in-school teacher training) to low-performing elementary schools in Santiago to improve students' academic achievement. I implemented an empirical strategy of difference-in-differences in the context of a quasi-experimental design to obtain an unbiased estimate of the causal effect of the TSFS Program on students' academic achievement. In particular, I obtained the first difference in academic achievement from two cohorts of students attending participating schools postprogram (treatment group) and preprogram intervention (control group), and the second difference, from students attending observably equivalent schools in nonprogram metropolitan areas during the same

TABLE 6. Program Impact 1 Year After Treatment: Fitted Multilevel Regression Models, Demonstrating the Relationship Between Students' Test Scores (Mathematics and Language) and Program Participation, Controlling for Selected School- and Student-Level Covariates

Variable	Model 1	Model 2	Model 3	Model 4
Outcome: mathematics				
Program Impact	4.42*	4.24*	3.89*	3.58†
Control for school-level covariates	NO	YES	NO	YES
Control for student-level covariates	NO	NO	YES	YES
R^2 between schools	0.7%	49.2%	22.3%	45.3%
R^2 within schools	0.2%	0.2%	3.4%	3.4%
N students	12,780	12,780	10,544	10,544
Outcome: language				
Program Impact	1.00	0.79	0.72	0.35
Control for school-level covariates	NO	YES	NO	YES
Control for student-level covariates	NO	NO	YES	YES
R^2 between-schools	4.1%	67.8%	15.5%	62.3%
R^2 within-schools	2.0%	2.0%	5.0%	5.0%
N students	12,869	12,780	10,584	10,584

Note. School $n = 138$. School-level covariates include Lang/Math 1999 test score, Eligible, Urban, Public, SES_1999, Enrollment_2001, Grade Retention_2001, Dropout_2001, SES_2006, SES_MID, SES_MIDLOW, SES_LOW, Enrollment_2006, Dropout_2006, Grade Retention_2006. Student-level covariates include: Income, Father's education, and Mother's education. POST-measure was 1 year after program implementation (SIMCE 2006). †$p < .10$. *$p < .05$.

Downloaded by [Universidad de Chile] at 11:54 03 February 2014

period of time (comparison group). I selected the comparison group using a propensity scores matching procedure. By subtracting the second difference from the first, I attempted to remove any historical trend in student achievement from my estimate of the causal impact of the program. Finally, I also controlled for several school- and student-level covariates.

I summarize the main findings of my study as follows. First, the TSFS Program had a statistically significant positive effect on fourth-grade students' academic achievement in both language and mathematics. Second, the estimated average program effect size was about 0.23 standard deviations in language and mathematics, and it was not sensitive

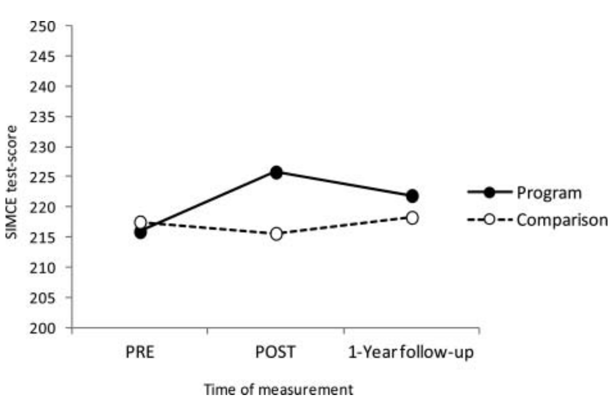

FIGURE 2. Estimated evolution of the average students' mathematics achievement in program and comparison schools. Outcome variable was SIMCE mathematics scores 1999 (PRE), 2005 (POST), and 2006 (1-year follow-up). Estimated SIMCE test scores for program and comparison groups, based on the corresponding multiple regression model (4) from Tables 4 and 6. See text for details.

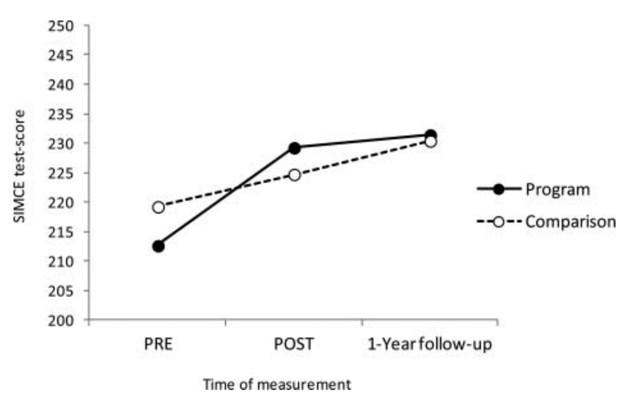

FIGURE 3. Estimated evolution of the average students' language achievement in program and comparison schools. Outcome variable: SIMCE language scores 1999 (PRE), 2005 (POST), and 2006 (1-year follow-up). Estimated SIMCE test scores for program and comparison groups, based on the corresponding multiple regression model (4) from Tables 4 and 6. See text for details.

Downloaded by [Universidad de Chile] at 11:54 03 February 2014

to control for relevant school- and student-level covariates. Third, although the program had a positive effect along the entire students' test-score distribution, there were larger effects for students situated in the middle part of that distribution (i.e., students scoring between deciles 2 and 7). The estimated program effect size measured at different deciles ranged from about 0.12 to about 0.30 standard deviations in SIMCE scores. This pattern was similar for language and mathematics academic achievement. Fourth, after the program intervention had ceased, program impact on test scores tended to decline rapidly: I found no program effect on language academic achievement in fourth graders evaluated 1 year after program implementation, and I estimated only a small positive effect on mathematics achievement (about one third of the original estimate). Finally, the TSFS Program also had a statistically significant negative effect on grade retention (which was also a program goal) with an estimated program impact of −0.49 standard deviations on grade retention rate among participant schools, a relatively medium to large effect size.

According to previous research on compensatory programs, the size of the estimated TSFS Program effect on students' academic achievement is noteworthy. It is about double the average effect size reported by evaluations of the U.S. Title I (Borman & D'Agostino, 2001), larger than the estimated impact of U.S. comprehensive school reform programs (Borman et al., 2003), and larger than the impact of the P-900 (Chay et al., 2005), the most relevant previous Chilean compensatory program in education. In addition, these findings indicate that the program impacted both mathematics and language academic achievement, in contrast to previous research that has found more consistent program effects on mathematics than on language achievement. Hence, from this point of view, the TSFS Program has been a comparatively successful educational policy.

What are the factors that could explain the effectiveness of the TSFS Program? My study did not consider this issue, but I can formulate some hypotheses based on additional evidence provided by recent qualitative research on the TSFS Program (Bellei, Raczynski, & Osses, 2010; Sotomayor & Dupriez, 2007). First, the kind of technical assistance that external consultants provided to schools was oriented to tackle practical real-work issues, by implementing in situ teacher training and mentoring, and different ways of coaching school principals. According to specialized literature, this type of professional development has been demonstrated to be more effective than the traditional approach based on courses taken at the universities (Elmore, 2004). Second, most of the external technical support teams implemented interventions designed around the findings of the school effectiveness and school improvement research: structured pedagogy, focus on classroom tasks, intensive use of materials guiding teaching practices, and directive and planned school change (Bellei Muñoz, Pérez, & Raczynski, 2004; Scheerens, 2000). Third, the TSFS Program incorporated some school accountability mechanisms, such as the use of external compulsory tests of students' academic achievement to set and evaluate school goals, and the threat of public sanctions for schools that failed to attain the defined goals at the end of the program.[15] Finally, the program intervention lasted 4 years, which is considered to be a minimum period of time for school improvement plans to be effective.[16] Future researchers should test the plausibility of these hypotheses.

A critical question for a pilot program such as the TSFS is whether its estimated impact on students' academic achievement can be replicated in different geographical zones or in a larger school population. Scaling up educational innovations requires subtle designs and careful implementations to overcome several cultural, organizational and institutional barriers present in the field of education (Elmore, 1996; Fullan, 1999). Presently, the Chilean government is expanding the use of external technical assistance to support school improvement plans focused on low-income students nationwide. For this purpose, the Ministry of Education organized a National Directory of Technical Assistance. The most challenging requirement for a successful expansion of the TSFS Program is the availability of high-quality consultant teams. In the TSFS Program, technical assistance was provided by skilled advisors, working mainly at the universities and prestigious research centers, who had institutional support to take part in a 4-year endeavor. This is not the case for most of the consultants participating in the recently created National Directory of Technical Assistance. Many of them work in small for-profit organizations or are independent professionals, have highly variable levels of experience and professional skills, and provide short-term advisory services within an autonomous market dynamic of demand–supply relationship between schools and technical assistance providers. Therefore, the effect of the TSFS Program estimated by this study may not necessarily be extended equally in this new educational policy setting. In this respect, the strategy followed by No Child Left Behind Act of 2001 (2002) of funding 16 Regional Comprehensive Centers to provide technical assistance to states and local districts, in order to improve their capacity to support school improvement processes, seems more promising to scaling up educational changes (Turnbull et al., 2011); nevertheless, its impact on students' academic achievement has not been evaluated yet.

On the other hand, conclusions as to the effectiveness of the TSFS Program should be made with caution given the observed disappearance of the TSFS effect on language, and its substantial decrease on mathematics 1 year after the conclusion of the program. This finding raises the question of the sustainability of the effect of the TSFS Program. In a recent study on this kind of external technical assistance programs in Chile (Bellei et al., 2010), some evidence of changes in classroom practices promoted by the programs was found, but there was no evidence of changes

Downloaded by [Universidad de Chile] at 11:54 03 February 2014

in teachers' capacities and enduring professional skills. This gap between monitored and autonomous school improvement may help to explain the pattern observed in this study.[17]

Additionally, there is a risk involved in using students' test scores for accountability purposes: Schools can concentrate their effort on the group of students whose test scores will be considered to evaluate their level of accomplishment with the defined goals. In fact, since the beginning of the intervention (2002), schools participating in the TSFS Program knew that they would be evaluated by the SIMCE results of the Grade 4 2005 cohort of students. This was also stated in the contract between the Ministry of Education and the external consultant teams. This kind of educational policy design can cause spurious increments on test scores (Koretz, 2008). While there is no direct evidence to support this hypothesis, it is highly consistent with the findings reported in this article. Notice that, as I also detected a similar pattern of diminishing program impact for science test scores (a subject matter that was not included in the official TSFS Program goals), my findings suggest a focus on the cohort more than a focus on the test effect. The issue of what kind of (intended and unintended) behaviors will be elicited by different school accountability designs has become a controversial educational policy debate in several countries, especially in the United States with the implementation of the No Child Left Behind Act. Present evidence suggests the need for prudence in implementing these sorts of policies because teachers, principals and students can respond in unpredictable and undesirable ways.

ACKNOWLEDGMENTS

I would like to thank Richard J. Murnane, John B. Willett, Fernando Reimers, and two anonymous reviewers for their valuable commentaries and suggestions.

NOTES

1. Launched in 1965, the program supplied additional funds to local educational administrations that served areas with defined concentrations of low-income families. The goal was to improve the education provided to disadvantaged students. Given the high level of autonomy of U.S. local authorities, Title I encompassed a great many heterogeneous educational interventions under this single funding umbrella. Technically, the current No Child Left Behind Act of 2001 (2002) is a reauthorization of this program.

2. The P-900 Program was the most important Chilean compensatory educational program implemented during the 1990s. It provided extra support (including teaching materials, teacher training, and extra-curricular activities) to low-performing primary schools that served low-income students. It was the most direct antecedent of the TSFS Program.

3. Certainly, according to studies such as Trends in International Mathematics and Science Study and Programme for International Student Assessment, U.S. and Chilean students—on average—perform very different in mathematics and language tests; nevertheless, in relative terms, these international comparisons show that both educational systems face similar challenges related to high levels of inefficiency and inequity. Additionally, both countries have educational systems in which local authorities are responsible for the administration of public schools, which creates a complex scenario for the design and implementation of nation-level educational policies; as a consequence, U.S. and Chilean policymakers are promoting the kind of programs here discussed in the context of standard based reforms.

4. The implementation of the program was limited to the Santiago metropolitan area for financial reasons, and as a pilot for other regions of the country.

5. The combined average score on these same tests was 250, nationally, and the standard deviation of the school mean was about 28.

6. I did not use SIMCE-2002 scores as the PRE measure, because the TSFS Program started in 2002. Because SIMCE was applied at the end of the academic year, the 2002 data may be affected by 1 year of treatment.

7. In the rest of the article, schools that entered the TSFS Program are referred to as the program schools and the schools that did not participate in the program are referred to as the comparison schools. Preprogram students are identified as the cohort of students who were tested before implementation and postprogram is the cohort of students tested after implementation. Finally, the group of students who attended program schools after program implementation is referred to as treated students.

8. Note that since program participation was decided at school level, I implemented the matching procedure at school level as well.

9. One of the 70 program schools did not have POST measures and was dropped from the sample.

10. As part of the sensitivity analysis, I also estimated program impact on grade retention rate (i.e., a school-level outcome).

11. Following Singer (1998), I estimated Model 1 and all other models using the MIXED procedure in SAS (Version 8.2, SAS Institute, Cary, NC) to account for the clustering of students within school, and to correctly compute standard errors.

12. To compute effect sizes, I divided the estimated regression coefficients by the standard deviation of the corresponding outcome in the program group: math ($SD = 47.57$) and language ($SD = 48.33$).

13. To compute effect sizes, I divided the estimated regression coefficient by the standard deviation of the corresponding outcome in the program group: Grade retention rate ($SD = 3.05$). Note that it is not possible to directly compare the reported program effect sizes on test scores and grade retention rates because the former is estimated by using a student-level measure as the outcome variable, while the latter is estimated by using a school-level outcome variable, which tends to overestimate program effect sizes. For comparison, when I replicated the estimates of program impact on test scores by using SIMCE school means as the outcome variable, the estimated program effect sizes increased to 0.83 standard deviations in mathematics and 0.57 standard deviations in language (from an original 0.23 standard deviations estimates in both subjects). This analysis is relevant to compare my results with previous research on compensatory programs in Chile; for example, Chay et al. (2005) used exclusively school-level data.

14. In analyses not reported here, I found a similar pattern for science test scores: the TSFS Program had a statistically significant positive effect on 2005 and 2006 students' achievement, but the estimated program effect size decreased from 0.29 to 0.16 standard deviations, respectively.

15. Actually, after the TSFS intervention, the government did not apply sanctions to schools that did not meet the stated program goals.

16. Borman et al. (2003) estimated that comprehensive school reforms needed about 5 years to significantly increase their impact on students' academic achievement. As a way of comparison, the mentioned Chilean P-900 program applied a 2-year cycle intervention

17. An alternative explanation is teachers' turnover, which is a common obstacle to school improvement programs in schools serving at-risk students. Unfortunately, although I have some impressionistic evidence about that issue provided by TSFS consultant teams, I have no data to test this hypothesis.

REFERENCES

Asesorías para el Desarrollo & Santiago Consultores. (2000). *Evaluación del Programa de Mejoramiento de la Calidad de las Escuelas Básicas de Sectores Pobres, P-900* [*Evaluation of the Program to Improve the Quality of Primary Schools Situated in Poor Areas, P-900*] (Unpublished Final Report). Santiago, Chile: Authors.

Bellei, C. (2003). ¿Ha tenido impacto la reforma educativa chilena? [Has Chilean educational reform had an impact?] In C. Cox (Ed.), *Políticas educacionales en el cambio de siglo: la reforma del sistema escolar de Chile* (pp. 125–209). Santiago, Chile: University of Chile Press.

Bellei, C., & Mena, I. (2000). The new challenge: quality and equity in education. In C. Toloza & E. Lahera (Eds.), *Chile in the nineties* (pp. 349–391). Palo Alto, CA: Stanford University.

Bellei, C., Muñoz, G., Pérez, L. M., & Raczynski, D. (2004). *¿Quién dijo que no se puede? Escuelas efectivas en sectores de pobreza* [*Who said it is not possible? Effective schools in poor areas in Chile*]. Santiago, Chile: UNICEF.
Bellei, C., Raczynski, D., & Osses, A. (2010). *Asistencia técnica educativa en Chile: ¿Aporte al mejoramiento escolar?* [*Educational technical assistance in Chile: A contribution to school improvement?*]. Santiago, Chile: Ocholibros Editores.
Borman, G., & D'Agostino, J. (2001). Title I and student achievement: a quantitative synthesis. In G. Borman, S. Stringfield, & R. Slavin (Eds.), *Compensatory education at the crossroads* (pp. 25–57). Mahwah, NJ: Erlbaum.
Borman, G., D'Agostino, J., Wong, K., & Hedges, L. (1998). The longitudinal achievement of Chapter 1 students: Preliminary evidence from the Prospects Study. *Journal of Education for Students Placed at Risk*, *3*, 363–399.
Borman, G., Hewes, G., Overman, L., & Brown, S. (2003). Comprehensive School Reform and achievement: A meta-analysis. *Review of Educational Research*, *73*, 125–230.
Bouveau, P. (2004). La discriminación positiva en el mundo: ¿una utopía pedagógica? [*The positive discrimination in the world: A pedagogical utopia?*]. In J. E. García-Huidobro (Ed.), *Políticas educativas y equidad* (pp. 49–57). Santiago, Chile: UNICEF.
Carter, L. (1984). The sustaining effects study of compensatory and elementary education. *Educational Researcher*, *13*(7), 4–13.
Chay, K., McEwan, P., & Urquiola, M. (2005). The central role of noise in evaluating interventions that use test scores to rank schools. *American Economic Review*, *95*, 1237–1258.
Cox, C. (2003). Las políticas educacionales de Chile en las últimas dos décadas del siglo XX [The Chilean educational policies in the last two decades of the 20th century]. In C. Cox (Ed.), *Políticas educacionales en el cambio de siglo: la reforma del sistema escolar de Chile* (pp. 19–113). Santiago, Chile: University of Chile Press.
Elmore, R. F. (1996). Getting to scale with good educational practice. *Harvard Educational Review*, *66*(1), 1–26.
Elmore, R. F. (2004). Bridging the gap between standards and achievement: the imperative for professional development in education. In R. F. Elmore (Ed.), *School reform from the inside out* (pp. 89–132). Cambridge, MA: Harvard Education Press.
Finnigan, K. & O'Day, J. (2003). *External support to schools on probation: Getting a leg up?* Philadelphia, PA: Consortium for Policy Research in Education.
Fullan, M. (1999). *Change forces: The sequel*. Philadelphia, PA: Falmer Press.
Gajardo, M. (2004). América latina: Políticas educativas de acción afirmativa. Situación, tendencias, perspectivas [Latin America: Educational policies of affirmative action: Situation, trends, perspectives]. In J. E. García-Huidobro (Ed.), *Políticas educativas y equidad* (pp. 101–118). Santiago, Chile: UNICEF.
Garcia-Huidobro, J. E., & Bellei, C. (2003). Desigualdad educativa en Chile [Educational inequality in Chile]. In R. Hevia (Ed), *La educación en Chile, hoy* (pp. 19–113). Santiago, Chile: Diego Portales University Press.
Koretz, D. (2008). *Measuring up: What educational testing really tell us*. Cambridge, MA: Harvard University Press.
Kosters, M., & Mast, B. (2003). *Closing the education achievement gap: Is Title I working?* Washington, DC: The AEI Press.
Laguarda, K. (2003). *State-sponsored technical assistance to low-performing schools: Strategies from nine states*. Washington, DC: Policy Studies Associates.
McDill, E., & Natriello, G. (1998). The effectiveness of the Title I Compensatory Education Programs: 1965–1997. *Journal of Education for Students Placed at Risk*, *3*, 317–335.
Meyer, B. D. (1995). Natural and quasi-experiments in economics. *Journal of Business & Economic Statistics*, *13*, 151–161.
Ministry of Education. (2006). *Asesoría a escuelas prioritarias* [*Technical support to priority schools*]. Santiago, Chile: Author.
No Child Left Behind Act of 2001, Pub. L. No. 107-110, § 115, Stat. 1425 (2002).
Organization for Economic Cooperation and Development. (2004). *Chile: Reviews of national policies for education*. Paris, France: Author.
Programa de las Naciones Unidas para el Desarrolo & Asesorías para el Desarrollo. (2004). *Plan de Asistencia Técnica para las Escuelas Críticas de la Región Metropolitana. Estudio de Seguimiento, informe final* [*Program of Technical Support to Failing Schools in Santiago Metropolitan Area: Monitoring study, final report*]. Santiago, Chile: Chilean Ministry of Education.
Puma, M. (1999, April). *The "prospects" study of educational growth and opportunity: Implications for policy and practice*. Paper presented at the Annual Meeting of the American Educational Research Association, Montreal, Quebec, Canada.
Puma, M., Karweit, N., Price, C., Ricciuti, A., Thompson, W., & Vaden-Kiernan, M. (1997). *Prospects: Final report on student outcomes*. Cambridge, MA: Abt Associates.
Reimers, F. (2000a). Conclusion: Can our knowledge change what low-income children learn? In F. Reimers (Ed.), *Unequal schools, unequal chances* (pp. 430–452). Cambridge, MA: David Rockefeller Center for Latin American Studies, Harvard University.
Reimers, F. (2000b). Educational opportunity and policy in Latin America. In F. Reimers (Ed.), *Unequal schools, unequal chances* (pp. 54–110). Cambridge, MA: David Rockefeller Center for Latin American Studies, Harvard University.
Rossi, P. H., Lipsey, M., & Freeman, H. (2003). *Evaluation: A systematic approach*. Thousand Oaks, CA: Sage.
Scheerens, J. (2000). *Improving school effectiveness*. Paris, France: UNESCO – IIEP.
Shadish, W. R., Cook, T. D., & Campbell, D. T. (2002). *Experimental and quasi-experimental designs for generalized causal inference*. Boston, MA: Houghton Mifflin.
Sheekey, A., Cymrot, D. J., & Fauntleroy, C. (2005). *Overview and synthesis of the Regional Advisory Committee Reports on Educational Challenges and Technical Assistance Needs*. Alexandria, VA: CNA Corporation.
Singer, J. D. (1998). Using SAS PROC MIXED to fit multilevel models, hierarchical models, and individual growth models. *Journal of Educational and Behavioral Statistics*, *24*, 323–355.
Sotomayor, C., & Dupriez, V. (2007). Desarrollar competencias docentes en la escuela: aprendizaje de una experiencia chilena de asesoría a escuelas de alta vulnerabilidad social y educativa [Develop teaching competences in the school: Lessons from a Chilean experience of technical support to socially and educationally disadvantaged schools]. *Les Cahiers de Recherche en Éducation et Formation*, *61*, 1–30.
Turnbull, B. J., White, R. N., Sinclair, E., Riley, D. L., & Pistorino, C. (2011). *National Evaluation of the Comprehensive Technical Assistance Centers: Final Report* (NCEE 2011–4031). Washington, DC: National Center for Education Evaluation and Regional Assistance, Institute of Education Sciences, U.S. Department of Education.
Weiss, C. (1998). *Evaluation. Methods for studying programs and policies*. Upper Saddle River, NJ: Prentice Hall.
Winkler, D. (2000). Educating the poor in Latin America and the Caribbean: Examples of compensatory education. In F. Reimers (Ed.), *Unequal schools, unequal chances* (pp. 112–133). Cambridge, MA: David Rockefeller Center for Latin American Studies, Harvard University.
Winship, C., & Morgan, S. (1999). The estimation of causal effects from observational data. *Annual Review of Sociology*, *25*, 659–706.
Wooldridge, J. M. (2002). *Econometric analysis of cross section and panel data*. Cambridge, MA: MIT Press.

AUTHOR NOTE

Cristián Bellei is Associate Researcher of the Center for Advanced Research in Education and Assistant Professor in the Sociology Department, both at the University of Chile. His main research areas are educational policy, school effectiveness, and school improvement; he has published extensively about quality and equity in Chilean education.

Downloaded by [Universidad de Chile] at 11:54 03 February 2014

Downloaded by [Universidad de Chile] at 11:54 03 February 2014

APPENDIX
Variable Definition and Quantile Regression Results

TABLE A1. Definitions of Variables

Variable	Definition
Outcome variables (student-level)	
MATHEMATICS	Mathematics SIMCE-2005 and SIMCE-2006 scores (IRT scale).
LANGUAGE	Language SIMCE-2005 and SIMCE-2006 scores (IRT scale).
Student level controls	
Grade repetition	Dummy variable indicating whether the student has previously been retained (retained = 1).
Mother's education	Years of education of the student's mother.
Father's education	Years of education of the student's father.
Income	An ordinal variable recording the student's family's monthly income, ranging from 0 (<100,000 Chilean Pesos) to 12 (>1,800,000 Chilean pesos).
School level controls	
Lang/Math_1999	Average Language/Mathematics SIMCE-1999 score.
Type of School	Dummy variable indicating the property status of the school. Public = 1; Private voucher = 0.
SES index	Percentage of socio-economically vulnerable students in the school (1999 and 2005). This is an official index used to distribute free-lunch.
School's students SES composition	Vector of three dichotomous variables recording the aggregate SES of students in the school, based on their aggregate family income, parental education, and percentage of vulnerable students in the school. Each variable is coded 1 to represent the named condition, 0 otherwise: (a) SES_LOW (low SES), (b) SES_MIDLOW (middle-low SES), and (c) SES_MID (middle SES).
Geographical location	Dummy variable for school location (Urban = 1, Rural = 0).
Enrollment	Number of students in the school (2001 and 2005).
Grade retention	Percentage of students retained in a grade (2001 and 2005).
Dropout rate	Percentage of students who dropped out of the school during the academic year (2001 and 2005).
Eligible	A dummy variable indicating a special school eligibility condition for the TSFS Program. Eligible = 1 if the school's average Language/Mathematics SIMCE-1999 score was less than 230 points and the school had been participating in a previous publicly funded school improvement program for more than 6 years; Eligible = 0 otherwise.

TABLE A2. Quantile Regression Coefficients (Figure 1)

	Outcome variable			
	Mathematics		Language	
Percentile	Model 1 without covariates	Model 2 with school-level covariates	Model 1 without covariates	Model 2 with school-level covariates
10th	6.81**	5.87**	8.32**	10.7***
20th	8.75**	9.7***	11.71***	13.04***
30th	12.64***	12.29***	14.13***	13.63***
40th	12.67***	11.09**	15.19***	14.79***
50th	14.1***	12.79***	12.36***	12.28***
60th	13.4***	11.99***	13.9***	13.08***
70th	10.19***	10.28***	13.03***	11.96***
80th	11.83***	10.72***	11.77***	9.75***
90th	8.27**	7.36**	8.48***	5.9*

Note. Estimated program effect at different levels of the students' academic achievement distribution. Difference-in-differences estimates of the program effect at the 10th, 20th, 30th, 40th, 50th, 60th, 70th, 80th, and 90th percentile of the students' test-score distributions in both mathematics and language. Outcome variable: SIMCE test scores 2005; student-level analysis. Coefficients were obtained from different quantile regression analyses. school-level covariates in models two include Lang/Math 1999 test score, Program, POST, Eligible, Urban, Public, SES_1999, Enrollment_2001, Grade Retention_2001, Dropout_2001, Enrollment_2005, Grade Retention_2005, Dropout_2005, SES_2005, SES_MID, SES_MIDLOW.
$^*p < .05$. $^{**}p < .01$. $^{***}p < .001$.