

# Can Schools Reduce the Indigenous Test Score Gap? Evidence from Chile

PATRICK J. MCEWAN

Wellesley College, USA

*Final version received March 2007*

**ABSTRACT** *In Chile, indigenous students obtain lower test scores, on average, than non-indigenous students. Between two cohorts of eighth-graders in the late 1990s, the test score gap declined by 0.1 to 0.2 standard deviations. An Oaxaca decomposition and related descriptive evidence suggest that the most plausible explanation is related to Chile's large-scale school reforms that were targeted at low-achieving schools and students. The paper evaluates and rules out alternate explanations such as relative improvements in indigenous socioeconomic status, and sorting of indigenous students between schools. The results highlight a potential lever for reducing earnings gaps between indigenous and nonindigenous adults.*

## I. Introduction

In 2002, almost 700,000 Chileans – 5 per cent of the population – identified themselves as indigenous. The majority of this group is Mapuche, one of the largest indigenous groups in South America (Grinevald, 1998; McEwan, 2004; Layton and Patrinos, 2006). A small literature documents that indigenous children in Chile perform worse than non-indigenous children on tests of academic achievement, with mean test scores that are approximately half a standard deviation lower (Herrera Lara, 1999; McEwan, 2004). Larger literature documents mean test score gaps between indigenous and non-indigenous students in primary schools throughout the western hemisphere.<sup>1</sup>

The causes of these gaps, particularly those related to primary schools, have important policy implications. Empirical evidence suggests that one-quarter of Chile's mean gap is explained by differences in the mean income and schooling of indigenous and non-indigenous parents (McEwan, 2004). However, more than half is explained by the differences in the quality of schools that are attended by indigenous and non-indigenous children. The remaining portion of the gap persists within schools, even among children with similar family characteristics.<sup>2</sup> In sum, cross-sectional evidence suggests that Chilean schools play some role in perpetuating the indigenous test score gap.

---

*Correspondence Address:* Patrick J. McEwan, Wellesley College, Department of Economics, 106 Central Street, Wellesley MA 02481, USA. Email: pmcewan@wellesley.edu

Since the resumption of democracy in 1990, education policy in Chile has emphasised direct investments in education quality by the Ministry of Education, often targeted at poor and low-achieving students (Cox, 1997, 2004; Delannoy, 2000; Garcia-Huidobro, 2000).<sup>3</sup> There is surprisingly little evidence on whether such programmes caused test scores to improve, and whether inequality declined as a result. Evaluations have found that some programmes, including investments in remedial tutoring and a lengthened school day, were successful in increasing mean school test scores (Chay et al., 2005; Valenzuela, 2005).

This paper bridges the two literatures, arguing that some primary school reforms contributed to a reduction of the indigenous test score gap across two cohorts of Chilean eighth graders in the late 1990s. The paper first confirms that indigenous students have lower mean test scores – by 0.4 to 0.5 standard deviations – than non-indigenous students in the 1997 cohort of eighth graders. In the 2000 cohort of eighth graders, the mean test score gap between indigenous and non-indigenous eighth graders declines by 0.2 standard deviations in Spanish, and 0.1 standard deviations in mathematics. There are two general explanations for the declining test score gap across cohorts.<sup>4</sup>

One possibility is that successive cohorts of indigenous families have experienced relative improvements in variables, like income or parental schooling, that affect children's test scores. While it is true that indigenous families have, on average, lower levels of income and parental schooling, these gaps did not change between 1997 and 2000. Using a decomposition procedure based on Cook and Evans (2000), the paper confirms that changing endowments of observed family variables cannot explain a declining mean test score gap.

In fact, the largest portion of the decline occurred within primary schools, among students with similar family characteristics, and the remaining balance occurred between primary schools. A range of explanations for this result is reviewed. The most plausible is related to the targeting and timing of Chile's large-scale school reforms that began in 1990 and rapidly expanded throughout the decade.<sup>5</sup> Some reforms were targeted towards poor and low-achieving schools. Even within such schools, relatively low-achieving students were more likely to participate in some programmes or may have reaped greater benefits from participating than higher-achieving children. The paper shows that, in 1997, indigenous students were more likely than non-indigenous students to attend a school treated by such reforms. By 2000, however, indigenous students were even more likely to have been exposed to such reforms. This divergence in the relative exposure of each group to school reform provides a plausible explanation for declining test score gaps. The conclusions discuss the broader policy implications for reducing inequality between indigenous and non-indigenous adults.

## **II. The Indigenous Population in Chile**

### *Population Size and Distribution*

The 2002 population census asked individuals to indicate whether they belong to one of eight indigenous groups.<sup>6</sup> Overall, 692,192 identified themselves as indigenous,

constituting 4.6 per cent of Chile's population. The vast majority identified themselves as Mapuche (87%), with the rest predominantly Aymara (7%) and Atacameño (3%). This population is not evenly dispersed throughout Chile's 13 regional territories (regions 1–12 are arrayed from north to south, while region 13 is dominated by Santiago and located in the centre of the country). The Mapuche population is concentrated in the southern region 9, and the country's capital in region 13, the latter due to migration.<sup>7</sup> In region 9, the epicenter of Mapuche culture, 24 per cent identify themselves as such, representing 29 per cent of the national Mapuche population. A smaller percentage (3%) identify themselves as Mapuche in region 13, but they account for 28 per cent of the nationwide population.

Other data confirm these patterns, with one exception.<sup>8</sup> In the 1992 census, 998,385 Chileans – 10 per cent of the population – identified themselves as indigenous (INE, 1993; McEwan, 2004).<sup>9</sup> The sources of this dramatic disagreement appear related to the much broader wording of the census question applied in 1992, suggesting that this paper's estimates estimate a lower bound to the true population.<sup>10</sup>

In other Latin American countries such as Bolivia and Guatemala, it is common to employ self-reported language competence as an additional indicator of indigenous status (McEwan, 2004; Layton and Patrinos, 2006; McEwan and Trowbridge, 2007). In Chile, the largest native language is Mapudungu, spoken by some Mapuche. It is cited by sociolinguists as one of the four largest native language groups in South America, with up to half a million speakers (see Grinevald, 1998 and the citations therein). Even so, these estimates are guesswork, and newer data indicate that the number may be substantially lower than surmised. The 2003 CASEN household survey included an additional question on self-reported language ability. Of the almost 700,000 that self-identify as indigenous, only 16 per cent report speaking and understanding an indigenous language, and 14 per cent report just understanding.<sup>11</sup> The endangered status of Mapudungu, though not the topic of this paper, is discussed further in a later section.

### *Population in Schools*

The indigenous population in the eighth grade is estimated with parent data collected in 1997 and 2000. Specifically, parent questionnaires ask whether a student's mother identifies herself as indigenous.<sup>12</sup> Overall, 5.1 per cent and 4.6 per cent of eighth graders have indigenous mothers in 1997 and 2000, respectively. Thus, the proportions in school are generally similar to those in the overall population.

In comparing the achievement of indigenous and non-indigenous students, a pitfall is sample selection. Suppose that indigenous students are less likely to enter school or complete primary school, and that non-attendees are lower-achieving, on average.<sup>13</sup> This would tend to understate the overall test score gap between indigenous and non-indigenous children, particularly in the upper grades. In fact, sample selection is not a primary concern in Chilean primary schools. In Chile, over 98 per cent of children ages seven to 13 – whether indigenous or non-indigenous – attend primary school, defined as grades 1–8.<sup>14</sup>

### III. Education Reform in the 1990s

In the 1980s, Chile's authoritarian government emphasised the decentralisation and privatisation of schooling.<sup>15</sup> In the 1990s, the democratic government initiated new programmes and reforms that were centrally financed and managed (Cox, 1997, 2004; Delannoy, 2000; Garcia-Huidobro, 2000). The following paragraphs describe three, in order of their implementation: (1) the 900 Schools Programme, or P-900; (2) the MECE programme (Equity and Quality Improvement in Education); and (3) the Full School Day reform.

P-900 was first implemented in the 1990 school year. The Ministry ranked primary schools according to their average test scores in 1988, and assigned a package of interventions to schools with 1988 scores that fell below region-specific eligibility cut-offs (Chay et al., 2005). The programme was applied in grades 1–4, and included minor infrastructure, instructional materials such as textbooks, in-service training for teachers in language and mathematics instruction, and intensive tutoring workshops for low-achieving children within the school. Some schools 'graduated' from P-900, typically after three years, while others were added throughout the decade, based upon test score performance and the subjective decisions of administrators.

In the 1992 school year, the Ministry initiated the MECE programme in primary schools. MECE contained multiple subcomponents, some available to all primary schools and some targeted according to various criteria. First, it distributed a wide range of instructional materials and infrastructure to all publicly-funded schools, including textbooks, classroom libraries, and computers. Second, the Ministry accepted applications for a diverse array of 'school improvement projects' from primary schools. It funded a portion of them with one-time grants, based upon the assessed potential of projects to contribute to student learning. Third, a component dubbed MECE-Rural provided special assistance to publicly-funded primary schools with only one or two teachers.

The third major reform, first implemented in the 1997 school year, subsidised an extended school day in publicly-funded schools. In grades three to six, annual instructional hours were increased by 232 hours, and, in grades seven and eight, by 145 hours (Cox, 2004). The reform was nominally available to all publicly-funded schools. In practice, it expanded first in schools that did not face a binding infrastructure constraint, since many schools were forced to move from a two-shift schedule to a single, longer shift. Many of the initial schools to participate were public schools, and those in rural areas (Raczynski et al., 2001; Valenzuela, 2005).

During this period, there were few special accommodations for indigenous populations, and indigenous self-identification was not used to target reforms. Herrera Lara (1999) notes that many rural Mapuche schools in southern regions, by virtue of their location and poverty, ended up participating in P-900 and MECE-Rural, and that some efforts were made to teach Spanish as a second language in P-900 schools. In the late 1990s, a specific programme of bilingual, intercultural education was started, but it was limited to small pilot initiatives carried out by regional universities (Gacitúa-Marió, 2002; Ministerio de Educación, 2005). The programme remained small in its coverage and its budget – especially compared to MECE – and it was only established as a national programme after 2000, the last year of data analysed in this paper (Ministerio de Educación, 2005).

#### IV. Data

##### *Sample*

Chile has administered a national census of student achievement (SIMCE) since 1988, with a different grade assessed in each year. For two reasons, this paper employs data from separate surveys of eighth grade cohorts in 1997 and 2000. First, detailed student-level data, including responses to parent surveys, are only available beginning in 1997.<sup>16</sup> Second, parents were asked to identify themselves as indigenous in these years, but not in later years. For example, fourth-grade surveys were administered in 1999 and 2002, but an indigenous variable is only available in 1999.<sup>17</sup>

The sample size in the original data files for 1997 and 2000 is 228,289 and 242,532, respectively, including all students with at least one test score in Spanish language or mathematics. Students were excluded if their parent did not report indigenous status, leading to the working sample sizes described in Table A1 (sample sizes are different across some variables because of missing observations for background variables).

##### *Variables*

The key dependent variables are test scores in Spanish language and mathematics. The test scores are standardised, within each subject and year, to a mean of zero and a standard deviation of one. Table A1 reports the descriptive statistics within each year. Family variables are derived from parent questionnaires. The essential variable is *Indigenous*, equal to one if a student's mother identifies herself as indigenous and zero otherwise. Other background variables include continuous measures of each parent's schooling (*Mother's schooling* and *Father's schooling*), and dummy variables indicating gender (*Female*) and four categories of monthly household income, measured in Chilean pesos.<sup>18</sup>

The data include a limited number of school variables. The percentage of indigenous students in each school is simply the school-level mean of *Indigenous*. *Average class size* is derived from administrative enrolment records. It is the total number of students enrolled in grades one to eight, divided by the total number of classroom sections. Finally, three dummy variables indicate three categories of public and private schools. *Public schools* are managed by more than 300 municipalities, but nationally-funded with per-student subsidies. *Private subsidised schools* are privately-managed, but receive similar per-student subsidies. Finally, *Private tuition schools* are privately-managed and receive no public funding.

#### V. Evidence of Declines in the Test Score Gap

Table 1 (panel A) describes the basic facts regarding mean test score gaps between indigenous and non-indigenous students. In the 1997 cohort of eighth graders, there are large and statistically significant gaps of  $-0.48$  and  $-0.41$  standard deviations in Spanish and mathematics, respectively.<sup>19</sup> In the 2000 cohort of eighth graders, the Spanish and mathematics gaps decline to  $-0.30$  and  $-0.32$ , respectively, also statistically significant. Overall, the mean difference between indigenous and non-indigenous test scores declines, between the two cohorts, by 0.18 standard deviations

Table 1. Variable means, by indigenous status and year

	1997 SIMCE			2000 SIMCE			Difference-in-differences
	Non-indigenous	Indigenous	Difference	Non-indigenous	Indigenous	Difference	
<i>Panel A: Test scores</i>							
Spanish test score	0.025	-0.457	-0.482**	0.014	-0.284	-0.298**	0.184**
Mathematics test score	0.021	-0.391	-0.412**	0.015	-0.310	-0.324**	0.087**
<i>Panel B: Student and family variables</i>							
Female	0.517	0.502	-0.016*	0.507	0.498	-0.009	0.007
Mother's schooling	9.814	7.421	-2.392**	10.210	7.761	-2.449**	-0.056
Father's schooling	10.242	7.856	-2.386**	10.737	8.312	-2.425**	-0.039
Monthly income							
<100,000 pesos	0.270	0.482	0.212**	0.286	0.514	0.228**	0.016
100,000-200,000 pesos	0.320	0.316	-0.003	0.337	0.317	-0.020**	-0.017
200,000-300,000 pesos	0.150	0.098	-0.052**	0.134	0.081	-0.053**	-0.001
>300,000 pesos	0.260	0.104	-0.156**	0.243	0.088	-0.155**	0.001
<i>Panel C: School variables</i>							
% indigenous in school	0.042	0.217	0.174**	0.037	0.234	0.197**	0.023
Average class size	35.413	33.397	-2.016**	35.844	33.194	-2.650**	-0.635*
Public school	0.572	0.694	0.122**	0.558	0.650	0.092**	-0.030
Private subsidised school	0.335	0.296	-0.039**	0.355	0.337	-0.018	0.021
Private paid school	0.094	0.010	-0.084**	0.087	0.013	-0.074**	0.010

Notes: \*\* and \* indicate statistical significance at 1 per cent and 5 per cent, respectively. See Table A1 for sample sizes in each year. See text for details of hypothesis tests.

in Spanish and 0.09 in mathematics. Both difference-in-differences estimates are statistically significant at one per cent.<sup>20</sup>

Whether the declines are economically significant is a more arbitrary judgment, although magnitudes are comparable to impact estimates of other education interventions in Chile. Quasi-experimental evaluations of the P-900 programme and the Full School Day programme report effect sizes of 0.1 to 0.2 standard deviations (Chay et al., 2005; Valenzuela, 2005). This paper's regression estimates, reported below, will suggest that increasing the mother's schooling by one standard deviation is associated with test score increases of around 0.1 standard deviation.

The prior estimates refer to differences in means, though test score gaps may be more or less evident at different points in the test score distribution.<sup>21</sup> Table 2 reports – in addition to mean differences between indigenous and non-indigenous test scores – the differences in each group's test scores at several percentiles of the respective test score distributions. The results suggest that test score gaps tend to increase at higher percentiles (except, in 1997, at the 90th). For Spanish test scores, the magnitude of declines in the gap is mostly stable across reported percentiles. For mathematics test scores, the magnitude of declines is largest at the 50th and 75th percentiles (though gaps decline at all percentiles).

Before attempting to explain the sources of declines in test score gaps, a caveat to the subsequent analysis is highlighted. The testing instruments applied in 1997 and 2000 are not composed of the same items. Because the difficulty of the tests could change across time, changes in the full sample means across years cannot be interpreted as aggregate changes in the achievement of eighth graders, and this paper does not attempt to do so. Similarly, it is not possible to determine whether changes in test score gaps are due to increased achievement among indigenous students, decreased achievement among non-indigenous students, or some combination thereof.

As a reasonable alternative, the author analyses how the relative test scores of indigenous and non-indigenous students – gauged by mean differences in z-scores –

**Table 2.** Indigenous test score gaps at various percentiles

	Spanish			Mathematics		
	Difference		Difference-in-differences	Difference		Difference-in-differences
	1997	2000		1997	2000	
<i>Percentile</i>						
10	-0.38	-0.15	0.22	-0.26	-0.23	0.03
25	-0.47	-0.25	0.22	-0.34	-0.31	0.03
50	-0.56	-0.34	0.22	-0.52	-0.34	0.17
75	-0.56	-0.37	0.19	-0.60	-0.38	0.23
90	-0.38	-0.39	-0.02	-0.43	-0.40	0.03
Mean	-0.48	-0.30	0.18	-0.41	-0.32	0.09

*Note:* The difference columns report, for each test score and year, the difference between a given percentile of the indigenous sample and the same percentile of the nonindigenous sample.

change over time. For this approach to be reasonable, it must be assumed that the 2000 test did not become relatively easier (or more difficult) for indigenous students. If it did, then declining test score gaps could be an artifact of test construction. A similar assumption is invoked in recent literature that compares black-white test score gaps in the US using non-equivalent tests (Hedges and Nowell, 1998, 1999), or analyses a longer panel of school-level mean test scores in Chile (Hsieh and Urquiola, 2006). The assumption is bolstered in this study's data by the apparent duplication of some testing items across the 1997 and 2000 tests.<sup>22</sup> Unfortunately, the assumption cannot be further validated without item-level data from the Chilean tests (which are not generally available to researchers), and so caution is warranted in assessing the following results, barring additional psychometric research.

## **VI. Explaining Declines in the Test Score Gap**

### *Families or Schools?*

There are two broad explanations for declines in the indigenous test score gap. The first is that indigenous families – though possessing smaller endowments of variables that improve test scores – narrowed this disadvantage between 1997 and 2000. To partly explore this hypothesis, Table 1 (panel B) compares non-indigenous and indigenous means of family variables. As with test scores, there are large and statistically significant gaps within each cross-section. In 1997, for example, the mean schooling of non-indigenous mothers and fathers is more than two years higher. In the same year, 27 per cent of non-indigenous families had monthly incomes below 100,000 pesos (expressed in nominal terms), compared with 46 per cent of indigenous families.

The substantial gaps may explain a portion of the gap within cross-sections, if these variables explain test scores. However, there is little change in the gaps between 1997 and 2000, and the difference-in-differences estimates for each family variable are not statistically distinguishable from zero. This suggests that declines in the test score gap will not be easily explained by shrinking gaps in observed socioeconomic status.

The second broad explanation for declines in the test score gap is that indigenous students experienced relative gains in schooling quality, compared to non-indigenous students. Table 1 (panel C) summarises the evidence on the few school characteristics reported in the data. In 1997, the average indigenous student attended a school in which 22 per cent of the other students were indigenous, while the average non-indigenous student attended a school with 4 per cent. This gap increased slightly in 2000 but the difference-in-differences is not statistically different from zero, suggesting that segregation by indigenous status has remained relatively stable.

Within each year, indigenous students attend schools with smaller than average class sizes, mainly because of schools disproportionate location in rural areas with few classrooms and smaller enrolments. Between 1997 and 2000, the gap expanded by less than one student. However, a recent quasi-experimental estimate of the effect of class size reduction on test scores in Chile suggests that such a change would

improve test scores by less than 0.01 standard deviations (McEwan and Urquiola, 2005).

Finally, indigenous students are less likely to attend private schools in both years, particularly private tuition-paying schools. However, the change in the gap between 1997 and 2000 is not statistically significant, suggesting that movements between the public and private sectors play a small role, if any, in explaining declines in the test score gap. In summary, the descriptive evidence does not suggest that declines in the mean test score gap are the result of either convergence in observed family variables that influence test scores, or in the relatively narrow set of observed school variables.

### *Empirical Approach*

The empirical approach is based upon the Oaxaca decomposition, commonly used to assess the determinants of mean wage gaps between two groups (Oaxaca, 1973).<sup>23</sup> Cook and Evans (2000) applied a variant of the method, also used in this paper, to black-white test score convergence in the US.<sup>24</sup> It must address two challenges presented by the Chilean data.

First, there are only a few school variables available, described in the last section. Second, it is likely that regression-based estimates of such variables' effects on test scores – even when controlling for family variables – will be biased by omitted variables. To provide one of the most common examples, it is likely that private school attendance is correlated with unobserved family variables such as wealth or parental behaviours. If omitted variables are also correlated with test scores, then the effect of private schools will be biased in ordinary least squares (OLS) regressions. This simple concern has spurred a large empirical literature in the US and Chile (Levin, 1998; McEwan, 2000, 2001). Similarly, other research on Chile has raised concerns about omitted variables bias in ordinary least squares estimates of the test score effects of class size reduction (McEwan and Urquiola, 2005), peer-effects (McEwan, 2003) and school reforms (Chay et al., 2005; Valenzuela, 2005).

One way of diminishing such concerns is to control for school fixed effects, facilitated by the availability of multiple student observations, both indigenous and non-indigenous, within schools. It begins with the estimation of OLS regressions within each year's sample (indicated by superscripts), as in:

$$T_{ij}^{97} = \beta_0^{97} + \beta_1^{97} \text{Indigenous}_{ij}^{97} + \beta_2^{97} X_{ij}^{97} + \mu_j^{97} + \varepsilon_{ij}^{97} \quad (1)$$

The test score ( $T$ ) of student  $i$  in school  $j$  is regressed on the dummy variable *Indigenous* and a vector of family-variables ( $X$ ) including parental schooling and household income.<sup>25</sup> The  $\beta$ 's are coefficients to be estimated, and each school is allowed to have a separate intercept via a series of fixed effects ( $\mu$ ). Finally,  $\varepsilon$  is a mean-zero error term. The standard errors are adjusted for arbitrary forms of heteroskedasticity and within-school clustering (Wooldridge, 2002).

The main advantage of the specification is that fixed effects control for all variables, observed and unobserved, that are constant across schools (though it prevents the estimation of separate coefficients for available school variables). Given

the coefficient estimates, the mean difference in indigenous and non-indigenous test scores within each cross-section can be re-written as:<sup>26</sup>

$$\underbrace{\Delta \bar{T}^{97}}_{\text{Test score difference}} = \underbrace{\hat{\beta}_1^{97}}_{\text{Within-school}} + \underbrace{\Delta \bar{X}^{97} \hat{\beta}_2^{97}}_{\text{Family endowments}} + \underbrace{\Delta \bar{\mu}^{97}}_{\text{School quality}} \quad (2)$$

where  $\Delta \bar{T}^{97} = \bar{T}_{Ind}^{97} - \bar{T}_{Non}^{97}$  and  $\Delta \bar{X}^{97} = \bar{X}_{Ind}^{97} - \bar{X}_{Non}^{97}$  (the overbars indicate variable means and the subscripts indicate the subsamples of indigenous and non-indigenous students). The first term on the right-hand side is the ‘unexplained’ portion of the test score gap – that is, the gap within each school that exists even among students with similar family characteristics. The second term is the portion of the gap that can be explained by differences in each group’s endowment of family variables such as parental schooling. The third term is the portion of the gap due to differences in the quality of schools, broadly construed, that are attended by indigenous and non-indigenous students. It can be further expressed as:

$$\Delta \bar{\mu}^{97} = \frac{1}{N_{Ind}} \sum_{j=1}^J \sum_{i=1}^{I_j} \text{Indigenous}_{ij}^{97} \hat{\mu}_j^{97} - \frac{1}{N_{Non}} \sum_{j=1}^J \sum_{i=1}^{I_j} (1 - \text{Indigenous}_{ij}^{97}) \hat{\mu}_j^{97} \quad (3)$$

where  $N$  indicates the total number of indigenous or non-indigenous students,  $J$  is the total number of schools, and  $I_j$  is the number of students within each school. The expression, though messy, has a straightforward interpretation. It is the average school fixed effect (that is, school quality) of each indigenous student, less the same average among non-indigenous students. An important caveat is that ‘school quality’ encompasses any variable that is constant across schools. This may include infrastructure, teacher quality and other instructional resources, but it also captures the school’s social context and peer-group characteristics.

The goal is to ascertain the causes of changes in the test score gap. Thus, Equation (2) can be successively applied to the 1997 and 2000 samples. The difference between the two can be written as:<sup>27</sup>

$$\underbrace{\Delta \bar{T}^{00} - \Delta \bar{T}^{97}}_{\text{Test score difference-in-differences}} = \underbrace{(\hat{\beta}_1^{00} - \hat{\beta}_1^{97})}_{\text{Within-school}} + \underbrace{(\Delta \bar{X}^{00} - \Delta \bar{X}^{97}) \hat{\beta}_2^{97}}_{\text{Family endowments}} + \underbrace{\Delta \bar{X}^{00} (\hat{\beta}_2^{00} - \hat{\beta}_2^{97})}_{\text{Family returns}} + \underbrace{(\Delta \bar{\mu}^{00} - \Delta \bar{\mu}^{97})}_{\text{School quality}} \quad (4)$$

The left-hand side indicates the test score difference-in-differences, already discussed in the previous section. The right-hand side decomposes it into four terms.

The first term estimates the change that occurs within schools, among students with similar family endowments of variables like parental schooling. The second term is the change in the gap that is due to changes in family endowments between indigenous and non-indigenous students. The third term measures whether changing returns to family variables (that is, the coefficients on family variables) are responsible for changes in the test score gap. Finally, the fourth term is the portion of changing gap that is attributable to changes in school quality.

### Regression Results

Table 3 reports estimates of Equation (1) for Spanish and mathematics test scores. The odd columns control only for *Indigenous*, and the coefficients replicate the previous results for test score gaps in 1997 and 2000. The regressions in even columns add controls for family variables and school fixed effects. The coefficients are consistent with previous analyses of student-level Chilean data (McEwan, 2001, 2003, 2004). There are large gender gaps in Spanish test scores of almost 0.2 standard deviations, favouring female students. Male students score higher, by a smaller margin, in mathematics. Parental schooling, especially of mothers, is strongly related to test scores. In 1997, for example, a one standard deviation increase in mother's schooling (3.8 years) increases Spanish test scores by 0.1 standard deviations. Relative to the lowest category of monthly household income, higher incomes are correlated with higher test scores. The causal interpretation of these coefficients is uncertain. Rather, the author anticipates that they are correlated with a broad range of unobserved indicators of the quantity and quality of educational inputs available in the home.

The coefficient on *Indigenous* is of particular interest. The inclusion of controls causes the coefficient estimates to decline substantially, suggesting that group differences in the endowments of family or school variables 'explain' the test score gap. Even so, the estimates in the 1997 sample suggest that a gap of almost 0.1 standard deviations persists within schools, even among observably similar students. By 2000, in contrast, estimates of the within-school gap are close to zero, or even positive. The counter-intuitive positive coefficient suggests that mean test score gaps within schools, among observationally similar children, have been reversed. The explanations of the result are explored at greater length in the following sections.

### Decomposition Results

Table 4 reports the results of the decomposition exercise for Spanish (panel A) and mathematics (panel B).<sup>28</sup> The left-most columns report the decomposition within the 1997 sample, based upon Equation (2).<sup>29</sup> Approximately one-quarter of the 1997 test score gaps can be explained by the different endowments of family variables across indigenous and non-indigenous students. A larger percentage, 55–60 per cent is explained by differing school quality. And, as noted above, the balance of the gap persists within schools among observably similar students.

The main objective, however, is not to decompose gaps within years, but to understand what caused the decline of the gap between 1997 and 2000. Thus, the right-most columns report the decomposition of the test score difference-in-differences in each subject, based upon Equation (4). Overall, the gap declines by 0.18 standard deviations in Spanish, and 0.09 in mathematics. The decomposition suggests that very little of this change is explained by changing endowments of family variables, consistent with previous descriptive evidence. Similarly, it cannot be easily explained by changes in the returns to family variables between years. In the case of mathematics, it even appears that changing returns to family characteristics explain an increase in the gap.

The largest portion of the decline in the test score gap occurred *within* primary schools, among students with similar endowments of family variables. In Spanish,

Table 3. Eighth-grade regressions

	Dependent variable: Spanish test score				Dependent variable: Mathematics test score			
	1997 SIMCE		2000 SIMCE		1997 SIMCE		2000 SIMCE	
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
<i>Indigenous</i>	-0.482** (0.017)	-0.098** (0.010)	-0.298** (0.016)	0.023* (0.011)	-0.412** (0.016)	-0.075** (0.010)	-0.324** (0.017)	0.012 (0.011)
<i>Female</i>	-	0.172** (0.005)	-	0.190** (0.005)	-	-0.066** (0.005)	-	-0.098** (0.005)
<i>Mother's schooling</i>	-	0.027** (0.001)	-	0.023** (0.001)	-	0.021** (0.001)	-	0.020** (0.001)
<i>Father's schooling</i>	-	0.014** (0.001)	-	0.017** (0.001)	-	0.011** (0.001)	-	0.016** (0.001)
<i>100,000-200,000 pesos</i>	-	0.077** (0.006)	-	0.102** (0.005)	-	0.039** (0.006)	-	0.094** (0.005)
<i>200,000-300,000 pesos</i>	-	0.103** (0.008)	-	0.137** (0.008)	-	0.058** (0.007)	-	0.134** (0.007)
<i>&gt; 300,000 pesos</i>	-	0.070** (0.008)	-	0.152** (0.008)	-	0.050** (0.008)	-	0.155** (0.008)
School fixed effects?	No	Yes	No	Yes	No	Yes	No	Yes
Adjusted R <sup>2</sup>	189,119	189,119	188,177	188,177	189,599	189,599	188,177	188,177
Number of students	189,119	189,119	188,177	188,177	189,599	189,599	188,177	188,177

Notes: \*\* and \* indicate statistical significance at 1 per cent and 5 per cent, respectively. Standard errors, adjusted for within-school clustering, are in parentheses. In addition to the listed variables, all regressions include a constant. Regressions in even columns include dummy variables that indicate missing values of *Female*, *Mother's schooling*, *Father's schooling*, and *Monthly income*.

**Table 4.** Decomposition results

	1997 SIMCE		2000 SIMCE		Difference	
<i>Panel A: Spanish test score</i>						
Raw difference	-0.48	100%	-0.30	100%	0.18	100%
Family variables						
Endowments	-0.12	25%	-0.13	42%	0.01	4%
Coefficients					-0.01	-6%
Within-school	-0.10	20%	0.02	-8%	0.12	65%
School quality	-0.26	55%	-0.19	65%	0.07	37%
<i>Panel B: Math test score</i>						
Raw difference	-0.41	100%	-0.32	100%	0.09	100%
Family variables						
Endowments	-0.09	22%	-0.11	35%	0.00	6%
Coefficients					-0.03	-30%
Within-school	-0.08	18%	0.01	-4%	0.09	100%
School quality	-0.25	60%	-0.22	69%	0.02	25%

65 per cent of the change occurred within schools and, in mathematics, it was 100 per cent (countervailed by the negative term on family returns). In each case, school quality also contributed to declines, but less so: 37 per cent in Spanish and 25 per cent in mathematics.

To summarise, the evidence suggests that convergence in endowments of family variables, at least those observed in the SIMCE data, plays little role in the declining gap. Instead, the declines occur mainly within schools among similar students or, to a lesser extent, between schools because of changes in school quality.

## VII. Did School Reform Reduce the Test Score Gap?

School reform might have affected the indigenous test score gap in two ways. First, it is possible that reforms were disproportionately targeted to schools that enroll indigenous students, and that this difference grew between the 1997 and 2000 cohorts. If school reform has a positive and homogenous causal effect on all students' test scores, then test score convergence would be reflected in the school quality component of the previous decomposition.

Yet, an even larger portion of test score convergence occurred within schools. For this to be the product of school reform, it must be the case that additional resources were either expressly targeted to indigenous students within schools, or that exposure to equivalent interventions had disproportionately large effects on the test scores of indigenous students.<sup>30</sup> The next two sections assess the likelihood of each possibility, focusing upon P-900 and the Full School Day, the two school reforms with the largest evidence base.

### *Differential Exposure to Reform Between Schools*

The 1997 cohort of eighth graders – this study's baseline – entered the first grade in 1990, while the 2000 cohort entered the first grade in 1993 (see Table 5).<sup>31</sup> P-900 began in 1990, and was targeted at grades one to four within participating schools.

**Table 5.** Expansion of P-900 and Full School Day

School year	Grade attended by eighth-grade cohort		Number of participating primary schools	
	1997	2000	P-900	Full school day
1990	1		969	–
1991	2		1,278	–
1992	3		1,123	–
1993	4	1	1,097	–
1994	5	2	1,060	–
1995	6	3	988	–
1996	7	4	900	–
1997	8	5	862	2,933
1998		6	893	3,521
1999		7	913	3,808
2000		8	909	4,267
Targeted schools?			Publicly-funded, Low-achieving	Publicly-funded, Rural*
Targeted grades?			1–4	3–8
Targeted students?			Low-achieving	–

*Note:* \*All publicly-funded schools are eligible but rural schools constituted a large number of initial participants.

*Source:* Cox (2004), Valenzuela (2005).

Thus, the 1997 cohort was potentially exposed to P-900 in the 1990–1993 school years, while the 2000 cohort was potentially exposed in the 1993–1996 school years.

There is reason to suppose that indigenous students were more likely to attend schools treated by the P-900 programme, and that this differential exposure grew over time. In the initial years of P-900 (1990–1992), programme allocation relied almost exclusively on school-level mean test scores. Despite the apparent objectivity of this approach, Chay et al. (2003, 2005) show that measurement error in school-level test scores led to some programme ‘misallocation’ in early years. Specifically, some relatively advantaged schools were inadvertently treated because they received transitorily low scores (just as some disadvantaged schools were not treated because of transitorily high scores). This occurred most frequently in very small schools that, because of sampling variation, were more likely to experience negative or positive shocks to their mean achievement. In subsequent years, the targeting relied less heavily on noisy test score measures, apparently incorporating the informed opinions of programme administrators (Chay et al., 2003, 2005). This implies that the quality of targeting towards disadvantaged schools – perhaps enrolling more indigenous students – improved over time.<sup>32</sup>

In the Full School Day reform, the logic is more straightforward. Given the timing of implementation, it could only have impacted the 1997 cohort of eighth graders in 1997 (see Table 5). In contrast, the 2000 cohort was potentially exposed to the reform in the 1997–2000 school years. Indigenous students were more likely to attend newly treated schools, since the early expansion of the programme was concentrated in rural and public schools attended disproportionately by such students (Raczynski et al., 2001; Valenzuela, 2005).

Notwithstanding these arguments, the changing relative exposure of indigenous students to school reform is an empirical question. Thus, the author combined the student-level SIMCE data from the 1997 and 2000 cross-sections with administrative records from each programme. For each student, a measure of potential exposure to each reform was constructed. The measures employ two key assumptions, given the lack of detailed panel data on each student's schooling career. First, it is assumed that members of each cohort did not repeat grades and entered school 'on-time'. Second, it was assumed that they attended the same school throughout their primary school career. This facilitates the calculation of two variables for each student: (1) a dummy variable indicating whether the student's school participated in at least one year of a reform; (2) a variable measuring the total number of years of potential exposure to school reform.

Table 6 reports means of each variable, for each reform, by indigenous status and year. Of non-indigenous students in the 1997 cross-section, for example, 20 per cent attended a school treated by P-900 for at least one year, compared with 36 per cent of indigenous students, a difference of 16 per cent. It is important to emphasise that indigenous students were not explicitly targeted. Rather, they were more likely to be treated because they were more likely to attend schools that happened to be treated. This difference is larger in 2000 (22%), again favouring indigenous students. The difference-in-differences is positive and statistically significant, suggesting a divergence in the potential exposure of indigenous and nonindigenous students to P-900.

Similarly, for the Full School Day, only 3 per cent of non-indigenous students participated in 1997, versus 11 per cent of non-indigenous students, a difference of 9 per cent. In the 2000 cohort, the difference grows to 18 per cent, again suggesting a divergence in the potential exposure of indigenous and non-indigenous students. The same pattern is evident when considering the total years of potential exposure to P-900 or the Full School Day reform.

Thus, the limited empirical evidence suggests that the potential exposure of indigenous students to school reform grew over time, relative to non-indigenous students. Two quasi-experimental evaluations suggest that exposure to each programme caused improvements in the mean achievement of students attending treated schools. Chay et al. (2005) applied a regression-discontinuity approach to the evaluation of P-900. Relying upon the test-score based programme allocation described in section III, it compared the fourth-grade test score gains between 1988–1992 of treated schools just below the eligibility cut-offs, to the gains of untreated schools just above the cut-off. By the end of 1992, treated schools test score gains were about 0.2 standard deviations higher.<sup>33</sup>

In an evaluation of the Full School Day, Valenzuela (2005) compared gains in mean fourth-grade test scores from 1999 and 2002, between treated and untreated schools. In Spanish, he found effects of 0.1 to 0.2 standard deviations among public schools, and 0.4 standard deviations among private subsidised schools. In mathematics, the effects are smaller or not statistically different from zero.

### *Differential Exposure to Reform Within Schools*

The majority of test score convergence occurred within primary schools, among indigenous and non-indigenous children with similar family characteristics. For school

**Table 6.** Potential exposure to school reform, by indigenous status and year

	1997 SIMCE			2000 SIMCE			Difference-in-differences
	Non-indigenous	Indigenous	Difference	Non-indigenous	Indigenous	Difference	
At least one year:							
<i>P-900</i>	0.198	0.356	0.158**	0.229	0.446	0.217**	0.059**
<i>Full School Day</i>	0.025	0.112	0.087**	0.079	0.262	0.183**	0.096**
Number of years:							
<i>P-900</i>	0.615	1.145	0.531**	0.629	1.227	0.598**	0.068
<i>Full School Day</i>	0.025	0.112	0.087**	0.209	0.792	0.583**	0.497**

Notes: \*\* (\*) indicates statistical significance at 1 per cent (5%). See Table A1 for sample sizes in each year. See text for details of hypothesis tests.

reform to be responsible, then it must either be the case that indigenous children were disproportionately targeted within schools, or that an intervention applied to all students within a school produced larger test score gains among indigenous students.

The most important component of P-900 was explicitly focused on low-achieving students within schools (Chay et al., 2005). Within participating schools, trained tutors worked with low-achieving students in grades one to four. The exact criteria used in student selection are not clearly described, nor are records on individual students participation maintained. As with school-level targeting, however, it is plausible that indigenous students were disproportionately targeted if the criteria were based on achievement or socioeconomic status.

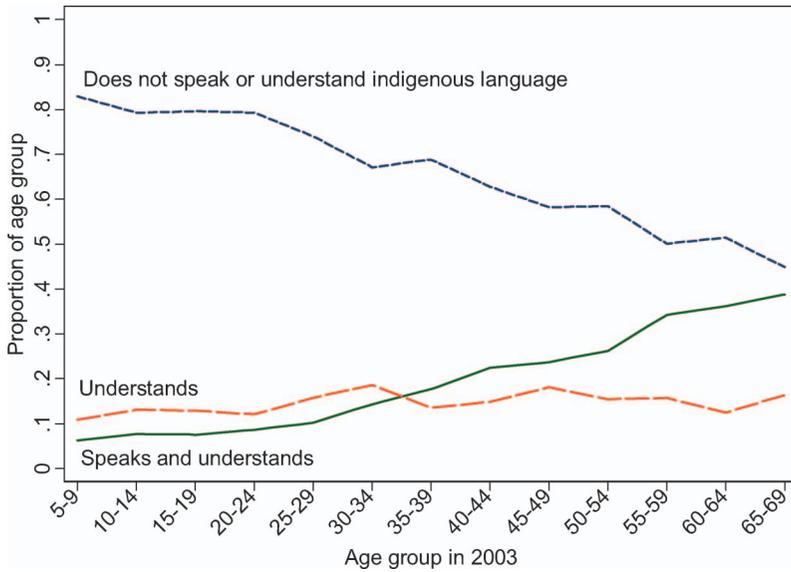
In the Full School Day programme, it is unlikely that indigenous students received more resources than non-indigenous students. However, the quasi-experimental evaluation suggests that lengthening the school day produced relatively larger test score gains among schools that enroll disproportionate numbers of disadvantaged and low-achieving students (Valenzuela, 2005). This evaluation, like others using Chilean data from the early 1990s, relies on data aggregated to the school level and cannot provide separate estimates for different groups of students. Thus, the case linking school reform to within-school declines in the test score gap remains circumstantial.

### *Alternate Explanations*

In light of uncertainties, five alternate explanations for the patterns described in the decomposition exercise are reviewed. First, declines in the test score gap within schools could be the result of convergence in unobserved family variables. As a plausible example, the relatively younger cohort of indigenous students in 2000 could be less likely to use native languages in the home, and more likely to use Spanish. In Bolivia and Guatemala, for example, younger cohorts report declining knowledge of indigenous languages (McEwan and Jiménez, 2002; Edwards and Winkler, 2004).<sup>34</sup> To the extent that this improves (Spanish) test scores over time, it could lead to reductions in test score gaps within schools.

Figure 1 reports data on native language ability from Chile's 2003 CASEN household survey, self-reported by survey respondents. The sample includes all individuals that identify themselves as indigenous. Within age group intervals, the figure reports the proportion of individuals that understand an indigenous language, speak and understand, and neither speak nor understand. The 1997 and 2000 cohorts of eighth graders would fall roughly within the survey cohort age of 16–19 years. As in other Latin American countries, the oldest cohorts are more likely to speak or understand a native language, but approximately half neither speak nor understand. Among the younger cohorts, between 5–24 years old, a fairly steady 80 per cent report no knowledge of a native language. Thus, while indigenous language ability has steadily declined across cohorts, it appears unlikely that small shifts in a three year period – and limited to 20 per cent of indigenous students – could account for the substantial within-school declines in Spanish test score gaps.

Second, it is possible that within-school declines are the result of sorting of indigenous and non-indigenous students across classrooms within schools. Suppose that indigenous and non-indigenous students in an older cohort were heavily segregated across classrooms within schools because of ability tracking or streaming.



**Figure 1.** Proportion of indigenous individuals with self-reported knowledge of indigenous languages. *Note:* Sample ( $N=22,309$ ) includes individuals that identify themselves as indigenous between the ages of 5–69. Proportions within each age group, defined in 5 year intervals, are weighted. *Source:* CASEN, 2003, and author’s calculations

Further suppose that a policy change diminished within-school segregation, leading to improved indigenous achievement (with a zero or small negative effect on non-indigenous students). This would appear in the previous decomposition as within-school test score convergence, though clearly unrelated to P-900 or other reforms. To assess this likelihood, the previous regressions with fixed effects for each classroom was re-estimated, facilitated by the availability of data on all eighth graders within a school. The revised decomposition is very similar to results with school fixed effects. It suggests that 59 per cent of Spanish test score convergence and 93 per cent of mathematics score convergence occurs within classrooms. Thus, sorting within schools is unlikely to be driving within-school declines in the test score gap.

Third, it is possible that declines are the product of students sorting between schools. That is, the school quality of students existing schools did not necessarily improve; instead, students in the 2000 cohort were more likely to choose high-quality schools. Two scenarios seem most plausible. In the first, indigenous families may have migrated from poor areas (such as region 9) to relatively richer communities, with higher-quality schools, such as Santiago in region 13. Indeed, such migration is responsible for the large number of Mapuche in Santiago. Yet, the indigenous population proportions in each region have remained steady across the 1997 and 2000 cohorts, which would be unlikely in the presence of extensive sorting across regions.

A related possibility is that sorting occurred between schools within local schooling markets, facilitated by Chile’s school choice provisions. That is, younger cohorts of families and children were more likely to choose higher-quality schools in their communities. However, evidence from Table 1 provided little evidence that the

proportion of indigenous families attending private schools has increased. Similarly, there was no evidence that students' exposure to other indigenous students has significantly changed over time, which might occur as a result of different school choices within local schooling markets

Fourth, it is possible that declines are the artifact of a shifting propensity among individuals to self-identify as indigenous. For this to be the case, indigenous families with relatively lower-achieving children must become less likely to identify as indigenous over a three year period, thus reducing the test score gap.<sup>35</sup> While it is plausible that indigenous self-identification could diminish in newer cohorts, it is not clear that such families would be among the lower-achieving indigenous families. It seems equally likely that such families would be drawn from the right tail of the indigenous achievement distribution, placing them above mean non-indigenous achievement, and leading to larger gaps.

Fifth, it is possible that school sample selection could bias the results. Some schools did not appear in both 1997 and 2000 samples, either because of closures and openings, or because SIMCE was not conducted in the school that year (though nominally universal, sometimes schools are excluded from sampling for administrative reasons). Thus, the regressions and decomposition with schools that appear in both rounds were re-estimated. The results do not substantially change.

## VIII. Conclusions

This paper has documented large differences in the mean achievement of indigenous and non-indigenous eighth graders in Chile. In three years, these mean differences declined by 0.2 standard deviations in Spanish and 0.1 in mathematics. Applying a decomposition procedure, it was shown that declines did not result from convergence in observed family characteristics. Rather, the majority of declines occurred within schools among students with similar socioeconomic status. A smaller portion, broadly attributed to school quality, occurred between schools.

The paper argues that the most plausible explanation for such declines in the test score gap is Chile's large-scale school reform in the 1990s. For two components, P-900 and the Full School Day, it was shown that indigenous students in 1997 were more likely to have been exposed to such reforms; this difference grew even larger in 2000. Furthermore, it was argued that at least one reform (P-900) targeted indigenous students within schools, and another (Full School Day) may have yielded a larger marginal impact among indigenous students. Though plausible, the case for school reform's role is circumstantial. Thus, I ruled out several alternative explanations, including student sorting. The strongest alternative explanation is that declines are an artifact of the different test instruments used in the two years, particularly if the 2000 test became relatively easier for indigenous students, but not others. It is important to replicate the facts described in this paper and probe their validity, should additional data become available, especially item-level test data.

Policies that affect test score gaps may have important implications for earnings inequality. Indigenous families in Chile – and, indeed, throughout Latin America – face relatively higher rates of poverty and extreme poverty than non-indigenous families. In 2000, 32.2 per cent of Chile's indigenous population fell below the poverty line, compared with 20.1 per cent of the non-indigenous population

(MIDEPLAN, 2002).<sup>36</sup> The Chilean gaps have persisted even in the wake of overall progress in poverty reduction since the late 1980s (Litchfield, 2002).

The sources of indigenous earnings inequalities have been little explored, but a small empirical literature, not including Chile, suggests that 50 to 60 per cent of the mean earnings disadvantage of indigenous groups is explained by observed differences in the quantity of human capital (that is, years of schooling).<sup>37</sup> A plausible explanation of the residual, in addition to labour market discrimination, is a difference in the quality of human capital obtained by groups (as might be proxied by test scores in Spanish or mathematics).<sup>38</sup> There is no research on whether test scores explain indigenous earnings gaps in Latin America, although US literature suggests that racial test score differences prior to labour market entry can explain a large proportion of the subsequent earnings gap (Neal and Johnson, 1996; Johnson and Neal, 1998). There is certainly a strong warrant for better exploring the links between schooling quality and earnings inequality in Chile, and in developing countries more generally.

Finally, this paper's findings merit an important caveat. The paper finds evidence of declines in Spanish and maths test score gaps, but not in other areas. In particular, education reforms did not necessarily promote fluency or literacy in native languages, such as Mapudungu, and may even have hastened their demise. In the early part of the 1990s, official Chilean education policy essentially ignored native languages. Bilingual education has belatedly assumed a small role in indigenous regions but it is dwarfed by resource investments in other reforms. There is a pressing need to credibly evaluate the impact of bilingual programmes – and more mature reforms – on achievement in Spanish and native languages.

### **Acknowledgements**

I am grateful to Gregory Elacqua and many officials of Chile's Ministry of Education for providing data. Gillette Hall, Joseph Shapiro, Miguel Urquiola, Emiliana Vegas and two anonymous referees provided helpful comments. The research received financial support from the World Bank, although the views expressed are only attributable to the author.

### **Notes**

1. The countries include Bolivia (Vera, 1998; McEwan, 2004), Canada (Ma and Klinger, 2000), Ecuador (Garcia Aracil and Winkler, 2004), Guatemala (Wu et al., 2003; Hernandez-Zavala et al., 2006; McEwan and Trowbridge, 2007), Mexico (Hernandez-Zavala et al., 2006), Peru (World Bank, 2001; Hernandez-Zavala et al., 2006), and the US (Freeman and Fox, 2005). See McEwan and Trowbridge (2007) for a summary.
2. Ethnographic evidence from region nine suggests that gaps within schools might be traced to low teacher expectations for indigenous students, and use of instructional materials more suited to non-indigenous, Spanish-speaking students (Herrera Lara, 1999).
3. The 1980s were characterised by public school decentralisation, privatisation and few direct interventions in school quality by the national Ministry of Education (McEwan and Carnoy, 2000; Hsieh and Urquiola, 2006).
4. This paper's task is similar to US literature that explores the causes for the decline of the black-white test score gap in the 1980s and its later stagnation or increase in the 1990s (Grissmer et al., 1998; Hedges and Nowell, 1998, 1999; Cook and Evans, 2000; Hanushek, 2001; Neal, 2005).

5. The 1980s reforms may have also played a role, albeit more indirect, in affecting test score gaps by removing constraints on school choice and perhaps allowing indigenous students to choose higher-quality schools. Nonetheless, both cohorts of students in this paper's data were exposed to a similar regime of school choice, and the evidence presented below does not suggest that re-sorting of indigenous students across schools can explain test score gaps.
6. The 2002 census asked: 'Do you belong to one of the following first or indigenous nations?' (author's translation). The mutually-exclusive options included Alacalufe (Kawashkar), Atacameño, Aimara, Colla, Mapuche, Qechhua, Rapa Nui, Yámana (Yagán), and none of the above.
7. Among Chile's smaller indigenous groups, the Aymara are concentrated in the far north (region one), near Bolivia, and the Atacameño in region two.
8. Chilean household surveys began collecting data on indigenous status in 1996. Using similar questions, the CASEN household surveys in 1996 and 2000 identified 4.5 per cent and 4.4 per cent of the population, respectively, as indigenous (MIDEPLAN, 2002). The CASEN household survey in 2003 found a slightly higher proportion indigenous (5.4%; author's calculations with CASEN 2003 data). The discrepancy could be due to sampling error or the use of imperfect weights derived from population projections based upon the 1992 census.
9. The 1992 census tabulations only report the indigenous population aged 14 or more (INE, 1993). Thus, the absolute number of the indigenous in 1992 would be even higher. In calculating the indigenous proportion, Gacitúa-Marió (2002) used the total population, ages zero and above, as the denominator, thus reporting a lower proportion.
10. The 1992 census asked: 'Do you feel that you belong to one of the following cultures?' (author's translation). Only three mutually exclusive options were provided (Mapuche, Aymara, and Rapanui).
11. The 2000 CASEN survey yields similar proportions (MIDEPLAN, 2002).
12. It asks: 'In Chile, the law acknowledges the existence of eight first nations or indigenous peoples; does the student's mother belong to one of them?' (author's translation). The mutually exclusive options are Atacameño, Aymara, Rapa-Nui, Quechua, Mapuche, Coya, Káwaskar and Yagán.
13. That is generally the case in countries without universal primary enrolments such as Guatemala (McEwan and Trowbridge, 2007).
14. The estimates are obtained from the 2003 CASEN household survey.
15. In 1980, the Ministry of Education decentralised the management of public schools to more than 300 municipalities (McEwan and Carnoy, 2000; Hsieh and Urquiola, 2006). It also began financing schools via a per-student subsidy based on monthly attendance. Private schools were eligible to accept such subsidies if they did not charge tuition. Parents were able to attend either public or private schools, presuming they were admitted. The reform precipitated a large-scale expansion of private schooling throughout the 1980s.
16. The 1997 data have also been used to analyse the effectiveness of private schools (McEwan, 2001) and the existence of peer-group effects (McEwan, 2003).
17. For an analysis of indigenous test score gaps in the 1999 cross-section, see McEwan (2004).
18. The original income categories in each year are more extensive and do not always overlap. They were collapsed to facilitate comparisons between the two survey years.
19. Within each year's sample, mean differences between indigenous and non-indigenous students are estimated by regressing test scores on *Indigenous*. Robust standard errors are adjusted for the clustering of students within schools.
20. Difference-in-differences are estimated by pooling the 1997 and 2000 samples and regressing students test score on *Indigenous*, a dummy variable indicating the 2000 sample, and the interaction between the two (where the last coefficient measures the difference-in-differences). Robust standard errors are adjusted for the clustering of students within year-by-school cells.
21. The US literature on black-white test score gaps has usually focused on mean test score differences. For an exception, see Clotfelter et al. (2006).
22. According to Cox (2004), it is possible to equate the 1997 and 2000 tests given anchor items duplicated on the 1997 and 2000 tests. He summarises such an exercise, and it shows that equated mean test scores in Spanish and mathematics are stable across time (implying similar mean achievement on anchor items).
23. For applications of the Oaxaca decomposition to the wage differences between indigenous and non-indigenous adults in Latin America, see Psacharopoulos and Patrinos (1994) and Hall and Patrinos (2006).

24. The present decomposition, and similar literature cited in footnotes 1 and 4, focuses on explaining mean test score gaps, and changes in mean gaps. Given the relative lack of empirical evidence in Chile or Latin America, this paper emphasises the application of well-known techniques for describing and decomposing mean test score differences. Nonetheless, the evidence in Table 2 suggested, primarily for math scores, that declines were more substantial at select percentiles of the indigenous and non-indigenous test score distributions. In future work, the author will explore this further with quantile regressions, though neither fixed-effects nor a Oaxaca-type decomposition can be applied without added methodological considerations (for example, see Koener, 2004; Machado and Mata, 2005).
25. Many control variables are missing values for some observations (see the sample sizes in Table A1). To avoid reducing sample size, the author arbitrarily recode them to their sample means, and further control for dummy variables equal to one for missing values, and zero otherwise (Krueger and Zhu, 2002). The coefficients on these additional variables, though not reported in the tables, are included in regressions and later decompositions. Conducting the analyses with a smaller sample that excludes all observations with missing data does not substantially alter the results.
26. Also see Cook and Evans (2000), McEwan (2004), and McEwan and Trowbridge (2007).
27. Note that the second term could also be weighted by the 2000 coefficients, and the third term by the 1997 mean differences, as in  $(\Delta\bar{X}^{00} - \Delta\bar{X}^{97})\hat{\beta}_2^{00} + \Delta\bar{X}^{97}(\hat{\beta}_2^{00} - \hat{\beta}_2^{97})$ , a well-known feature of an Oaxaca-type decomposition. All subsequent estimates were repeated both ways, but it does not substantially affect the results.
28. A potential disadvantage of Oaxaca-type decompositions based on linear regressions is that observations in the indigenous and non-indigenous groups may not fall within a region of common support (see especially Nopo, 2004 and its references). That is, some indigenous children may be observationally quite dissimilar from non-indigenous children, and vice-versa, and the decomposition implicitly relies on projections of outcomes outside the observed range for such students. To assess the sensitivity of results, probits within each cross-sectional sample were estimated, regressing the indigenous dummy variable on the full set of family and school characteristics listed in Table A1. Propensity scores for each student were then calculated. In 1997 (2000), the region of common support includes 98.0% (97.7%) of non-indigenous students and 97.9% (98.6%) of indigenous students. The author then dropped observations outside of the region of common support and re-estimated the fixed-effects regressions in Table 3. The estimates and resulting decomposition results were not substantively different from the full sample estimates.
29. McEwan (2004) carried out a similar exercise; the results are similar, but not identical because of small sample differences.
30. For a similar discussion in the context of black-white test score convergence in the US, see Hanushek (2001).
31. Due to grade repetition, some eighth grade students may have spent more than eight years in school and, therefore, entered earlier than 1990 or 1993. The issue is less severe in Chile than other Latin American systems because of generally low rates of grade repetition, less than 3 per cent in the first grade (McEwan and Shapiro, 2008).
32. Declines in the test score gap between schools could have occurred even if indigenous students in the 2000 cohort were not more likely to be targeted by P-900 than students in the 1997 cohort. Anecdotal and empirical evidence that the programme became more effective over time, probably because programme interventions were added in the 1992 school year (Chay et al., 2005). Thus, the 2000 cohort may have been exposed to a relatively more effective intervention.
33. As Chay et al. (2005) note, there is no evidence on the relative cost-effectiveness of P-900 and other reforms in improving student achievement.
34. Despite this evidence, there is a surprising lack of descriptive evidence on how knowledge of native languages has shifted over time, and how such shifts may be related to children's school performance.
35. It may also be the case that non-indigenous families with high-achieving children become more likely to identify as indigenous over time. This seems implausible, given that indigenous identify is stigmatised in Chile, revealed in attitude surveys (see Gacitúa-Marió, 2002 and the citations therein).
36. The indigenous are also more likely to fall below the extreme poverty line (11% of the indigenous population, compared with 5.4% of the non-indigenous population). For similar data on other Latin American countries, see Psacharopoulos and Patrinos (1994), de Ferranti et al. (2004), World Bank (2004), Hall and Patrinos (2006).

37. In particular, see de Ferranti et al. (2004: 94) and Psacharopoulos and Patrinos (1994: 214). In each case, the results are obtained from an application of the Oaxaca decomposition to estimates from Mincer earnings regressions estimated separately among indigenous and nonindigenous adults.
38. It is also plausible that the gap reflects differences in Spanish language skills – as might be captured by test scores – an interpretation forwarded in the Bolivian context by Chiswick et al. (2000).

## References

- CASEN (2003) Encuesta de Caracterización Socioeconómica [Socioeconomic Characterization Survey], Ministerio de Planificación, Santiago, Chile.
- Chay, K.Y., McEwan, P.J. and Urquiola, M. (2003) The central role of noise in evaluating interventions that use test scores to rank schools. Working Paper No. 10118, National Bureau of Economic Research, Cambridge, MA.
- Chay, K.Y., McEwan, P.J. and Urquiola, M. (2005) The central role of noise in evaluating interventions that use test scores to rank schools. *American Economic Review*, 95(4), pp. 1237–1258.
- Chiswick, B.R., Patrinos, H.A. and Hurst, M.E. (2000) Indigenous language skills and the labor market in a developing economy: Bolivia. *Economic Development and Cultural Change*, 48(2), pp. 349–367.
- Clotfelter, C.T., Ladd, H.F. and Vigdor, J.L. (2006) The academic achievement gap in grades 3 to 8. Working Paper No. 12207, National Bureau of Economic Research, Cambridge, MA.
- Cook, M.D. and Evans, W.N. (2000) Families or schools? Explaining the convergence in white and black academic performance. *Journal of Labor Economics*, 18(4), pp. 729–754.
- Cox, C. (1997) La reforma de la educación chilena: contexto, contenidos, implementación. Documentos de trabajo 8, Programa de Promoción de la Reforma Educativa en América Latina, Washington, DC.
- Cox, C. (2004) Innovation and reform to improve the quality of primary education: Chile. Unpublished manuscript, Ministry of Education, Santiago.
- de Ferranti, D., Perry, G.E. Ferreira, F.H.G. and Walton, M. (2004) *Inequality in Latin America: Breaking with History* (Washington, DC: World Bank).
- Delannoy, F. (2000) *Education Reforms in Chile, 1990–1998: A Lesson in Pragmatism* (Washington, DC: World Bank).
- Edwards, J.H.Y. and Winkler, D.R. (2004) Capital humano, globalización y asimilación cultural: un estudio aplicado a los Mayas de Guatemala, in: D.R. Winkler and S. Cueto (eds), *Etnicidad, Raza, Género y Educación en América Latina* (Santiago: PREAL), pp. 133–177.
- Freeman, C. and Fox, M. (2005) Status and trends in the education of American Indians and Alaska Natives. National Center for Education Statistics, Report 2005-108, US Department of Education.
- Gacitúa-Marió, E. (2002) Indigenous peoples in Chile: current situation and policy issues, in: *Chile's High Growth Economy: Poverty and Income Distribution, 1987–1998* (Washington, DC: World Bank), pp. 214–238.
- García Aracil, A. and Winkler, D. (2004) Educación y etnicidad en Ecuador, in: D.R. Winkler and S. Cueto (eds), *Etnicidad, Raza, Género y Educación en América Latina* (Santiago: PREAL), pp. 55–92.
- García-Huidobro, J.E. (2000) Educational policies and equity in Chile, in: F. Reimers (ed.), *Unequal Schools, Unequal Chances: The Challenges to Equal Opportunity in the Americas* (Cambridge, MA: Harvard University, David Rockefeller Center for Latin American Studies), pp. 161–181.
- Grinevald, C. (1998) Language endangerment in South America: a programmatic approach, in: L.A. Grenoble and L.J. Whaley (eds) *Endangered Languages: Current Issues and Future Prospects* (Cambridge: Cambridge University Press), pp. 124–159.
- Grissmer, D., Flanagan, A. and Williamson, S. (1998) Why did the black-white score gap narrow in the 1970s and 1980s?, in: C. Jencks and M. Phillips (eds), *The Black-White Test Score Gap* (Washington, DC: Brookings Institution Press), pp. 182–226.
- Hall, G. and Patrinos, H.A. (eds) (2006) *Indigenous People, Poverty and Human Development in Latin America, 1994–2004* (New York, NY: Palgrave Macmillan).
- Hanushek, E.A. (2001) Black-white achievement differences and governmental interventions. *American Economic Review*, 91(2), pp. 24–28.
- Hedges, L.V. and Nowell, A. (1998) Black-white test score convergence since 1965, in: C. Jencks and M. Phillips (eds), *The Black-White Test score Gap* (Washington, DC: Brookings Institution Press), pp. 149–181.
- Hedges, L. and Nowell, A. (1999) Changes in the black-white gap in achievement test scores. *Sociology of Education*, 72(2), pp. 111–135.

- Hernandez-Zavala, M., Patrinos, H.A. Sakellariou, C. and Shapiro, J. (2006) Quality of schooling and quality of schools for indigenous students in Guatemala, Mexico and Peru. Unpublished manuscript, World Bank.
- Herrera Lara, R.H. (1999) Cultural images, education and domination: the case of the Mapuche of southern Chile. Doctoral dissertation, University of Illinois at Urbana-Champaign.
- Hsieh, C.-T. and Urquiola, M. (2006) The effects of generalized school choice on achievement and stratification: evidence from Chile's school voucher program. *Journal of Public Economics*, 90(8–9), pp. 1477–1503.
- INE. See Instituto Nacional de Estadística.
- Instituto Nacional de Estadística (1993) Resultados oficiales, censo de población. Instituto Nacional de Estadística, Santiago.
- Johnson, W.R. and Neal, D. (1998) Basic skills and the black-white earnings gap, in: C. Jencks and M. Phillips (eds), *The Black-White Test Score Gap* (Washington, DC: The Brookings Institution Press), pp. 480–497.
- Koenker, R. (2004) Quantile regression for longitudinal data. *Journal of Multivariate Analysis*, 91(1), pp. 74–89.
- Krueger, A.B. and Zhu, P. (2002) Another look at the New York City school voucher experiment. Working Paper No. 9418, National Bureau of Economic Research, Cambridge, MA.
- Layton, H.M. and Patrinos, H.A. (2006) Estimating the number of indigenous people in Latin America, in: G. Hall and H.A. Patrinos (eds), *Indigenous People, Poverty and Human Development in Latin America, 1994–2004* (New York, NY: Palgrave Macmillan).
- Levin, H.M. (1998) Educational vouchers: effectiveness, choice, and costs. *Journal of Policy Analysis and Management*, 17(3), pp. 373–391.
- Litchfield, J. (2002) Updated income distribution and poverty measures for Chile: 1987–98, in: *Chile's High Growth Economy: Poverty and Income Distribution, 1987–1998* (Washington, DC: The World Bank), pp. 47–72.
- Ma, X. and Klinger, D.A. (2000) Hierarchical linear modeling of student and school effects on academic achievement. *Canadian Journal of Education*, 25(1), pp. 41–55.
- Machado, J.A.F. and Mata, J. (2005) Counterfactual decomposition of changes in wage distributions using quantile regression. *Journal of Applied Econometrics*, 20(4), pp. 445–465.
- McEwan, P.J. (2000) The potential impact of large-scale voucher programs. *Review of Educational Research*, 70(2), pp. 103–149.
- McEwan, P.J. (2001) The effectiveness of public, Catholic, and non-religious private schools in Chile's voucher system. *Educational Economics*, 9(2), pp. 103–128.
- McEwan, P.J. (2003) Peer effects on student achievement: evidence from Chile. *Economics of Education Review*, 22(2), pp. 131–141.
- McEwan, P.J. (2004) The indigenous test score gap in Bolivia and Chile. *Economic Development and Cultural Change*, 53(1), pp. 157–190.
- McEwan, P.J. and Carnoy, M. (2000) The effectiveness and efficiency of private schools in Chile's voucher system. *Educational Evaluation and Policy Analysis*, 22(3), pp. 213–239.
- McEwan, P.J. and Jiménez, W. (2002) Indigenous students in Bolivian primary schools: patterns and determinants of inequities. Girls' Education Working Papers, World Bank, Washington, DC.
- McEwan, P.J. and Shapiro, J.S. (2008) The benefits of delayed primary school enrollment: discontinuity estimates using exact birth dates. *Journal of Human Resources*, 43(1), pp. 1–29.
- McEwan, P.J. and Trowbridge, M. (2007) The achievement of indigenous students in Guatemalan primary schools. *International Journal of Educational Development*, 27(1), pp. 61–76.
- McEwan, P.J. and Urquiola, M. (2005) Precise sorting around cutoffs in the regression-discontinuity design: evidence from class size reduction. Unpublished manuscript, Wellesley College and Columbia University.
- MIDEPLAN. See Ministerio de Planificación y Cooperación.
- Ministerio de Educación (2005) Orientaciones programa de educación intercultural bilingüe. Ministerio de Educación, Santiago.
- Ministerio de Planificación y Cooperación (2002) Análisis de la VIII Encuesta de Caracterización Socioeconómica Nacional (CASEN 2000). Documento No. 14, Etnias y Pobreza en Chile, Ministerio de Planificación y Cooperación, Santiago.

- Neal, D. (2005) Why has black-white skill convergence stopped? Working Paper No. 11090, National Bureau of Economic Research, Cambridge, MA.
- Neal, D. and Johnson, W.R. (1996) The role of premarket factors in black-white wage differences. *Journal of Political Economy*, 104(5), pp. 869–895.
- Ñopo, H. (2004) Matching as a tool to decompose wage gaps. Discussion Paper No. 981, IZA, Bonn, Germany.
- Oaxaca, R. (1973) Male-female wage differentials in urban labor markets. *International Economic Review*, 14(3), pp. 693–709.
- Psacharopoulos, G. and Patrinos, H.A. (eds) (1994) *Indigenous People and Poverty in Latin America: An Empirical Analysis* (Washington, DC: World Bank).
- Raczynski, D., Ruz P., M.A., Madrid V., A., Pavez G., A. and Quiroga M., P. (2001) *Estudio de Evaluación Jornada Escolar Completa* (Santiago: Pontificia Universidad Católica de Chile).
- Valenzuela, J.P. (2005) Partial evaluation of a big reform in the Chilean education system: From a half day to a full day schooling. Unpublished manuscript, University of Michigan.
- Vera, M. (1998) Bilingüismo y rendimiento escolar en Bolivia. MA thesis, ILADES-Georgetown.
- Wooldridge, J. (2002) *Econometric Analysis of Cross Section and Panel Data* (Cambridge, MA: MIT Press).
- World Bank (2001) *Peruvian Education at a Crossroads: Challenges and Opportunities* (Washington, DC: World Bank).
- World Bank (2004) *Poverty in Guatemala* (Washington, DC: World Bank).
- Wu, K.B. Goldschmidt, P. and Hara, M. (2003) The determinants of student achievement in primary education, in: *Guatemala: Equity and Student Achievement in Primary Education (Vol. 2: Background Papers)* (Washington, DC: World Bank).

**Table A1.** Descriptive statistics

	1997 SIMCE		2000 SIMCE	
	Mean (S.D.)	Number of students	Mean (S.D.)	Number of students
Spanish test score	0.00 (1.00)	189,119	0.00 (1.00)	188,177
Math test score	0.00 (1.00)	189,599	0.00 (1.00)	188,177
Indigenous	0.05	194,359	0.05	188,177
Female	0.52	191,449	0.51	188,176
Mother's schooling	9.69 (3.80)	190,176	10.10 (3.99)	175,140
Father's schooling	10.12 (4.08)	187,788	10.63 (4.26)	167,653
Monthly household income				
< 100,000	0.28	188,913	0.30	182,672
100,000–200,000	0.32	188,913	0.34	182,672
200,000–300,000	0.15	188,913	0.13	182,672
> 300,000	0.25	188,913	0.24	182,672
% indigenous in school	0.05 (0.09)	194,359	0.05 (0.09)	188,177
Average class size	35.31 (7.02)	190,581	35.72 (7.02)	185,939
Public school	0.58	194,359	0.56	188,177
Private subsidised school	0.33	194,359	0.36	188,177
Private paid school	0.09	194,359	0.08	188,177