

Compensatory Education for Disadvantaged Mexican Students: An Impact Evaluation Using Propensity Score Matching

Joseph Shapiro and Jorge Moreno Trevino¹

Abstract. This paper uses propensity score matching to evaluate the effectiveness of CONAFE, a compensatory education program in Mexico, in improving student test scores and lowering repetition and failure rates. We find that CONAFE is most effective in improving primary school math learning and secondary school Spanish learning. *Telesecundaria* education and bilingual education for indigenous students are both shown to improve student achievement. CONAFE also lowers primary school repetition and failure rates. We conclude that this compensatory education program can effectively improve short-term learning results for disadvantaged students, but that improvement varies by the subject of instruction and the demographic of student taught.

JEL Codes: I20, I21, I28

World Bank Policy Research Working Paper 3334, June 2004

The Policy Research Working Paper Series disseminates the findings of work in progress to encourage the exchange of ideas about development issues. An objective of the series is to get the findings out quickly, even if the presentations are less than fully polished. The papers carry the names of the authors and should be cited accordingly. The findings, interpretations, and conclusions expressed in this paper are entirely those of the authors. They do not necessarily represent the view of the World Bank, its Executive Directors, or the countries they represent. Policy Research Working Papers are available online at <http://econ.worldbank.org>.

¹ World Bank (IBRD). Email: jshapiro2@worldbank.org, jmoreno@colmex.mx

This research benefitted from the insightful direction and advice of Harry Anthony Patrions. This paper also benefitted from comments and advice from Ariel Fiszbein, Gladys Lopez-Acevedo, Claudia Macias, Victor Paqueo, Anna Sant'anna and Emiliana Vegas (World Bank), participants in a November 2003 World Bank LCSHD discussion seminar, María Ortíz, Joaquín Quesada, Edgar Andrade, and Martha Palacios (Secretaría de Educación Pública - Mexico), and Yessica Montiel and Felipe Cuéllar (CONAFE).

Introduction

The President of Mexico, Vicente Fox, proclaimed in 2001 that his top education priority was the provision of equal and high-quality education to all Mexicans—a sentiment that reaffirmed the commitments of earlier Mexican administrations (Secretaría de Educación Pública 2001). But effectively and cost-effectively educating all citizens is difficult in a geographically disperse and culturally heterogeneous country like Mexico. How should Mexico educate the type of students who speak no Spanish, live in villages inaccessible by roads, or come from families that cannot afford school uniforms?

A similar question should concern education policymakers in most poor- and middle-income countries. High-quality basic education is necessary to end the transmission of poverty from one disadvantaged generation to the next. Health and infrastructure improvements can build a framework for every person to live a life free of poverty. But if a country's most disadvantaged students do not receive a high-quality education, those students will be largely unable to escape the intractable and abject poverty that characterizes too many disadvantaged communities.

Mexico began to address this challenge as early as 1971 by creating the National Council of Education Promotion (CONAFE), a division of Mexico's Secretariat of Public Education (SEP). CONAFE provides extra resources to schools that enroll disadvantaged students. CONAFE's compensatory education programs now support about four million students in preschool and primary education, and about 300,000 students in *telesecundaria* education (secondary education delivered via television to remote communities).²

The present document studies the effectiveness of CONAFE in improving basic education for Mexico's most disadvantaged students. While our analysis should particularly

² Current numbers based on 1998 projections.

interest researchers and policymakers who study Mexican education, it should also concern education researchers generally as a case study of how a country can effectively improve the learning outcomes of disadvantaged students. One might think that a country's worst-performing schools lack the necessary bureaucratic and physical infrastructure to effectively use extra resources, so providing extra resources to severely disadvantaged schools would cause little impact on student learning outcomes. This paper profiles one case where extra resources appear to have measurable impact on student learning, though the unique features of the CONAFE program in Mexico may imply that not all compensatory education programs will have similar impact.

Background

CONAFE compensates for the early disadvantages of some students by providing extra resources to the schools enrolling those students. In this regard, it is a program of compensatory education, and its model resembles compensatory education elsewhere. An array of economic literature has attempted to evaluate compensatory education outside Mexico.

Among the largest compensatory education programs is Chapter 1 (formerly Title 1) in the U.S., which in 1991 allocated \$4.3 billion to schools enrolling low-income students. A variety of evaluations in the 1980s found that Chapter 1 effectively increased test scores over a 1-year period, but that scores fell in the summer or year following the investment of Chapter 1 (Slavin 1989).

More recent research has focused on the Head Start program in the U.S., which provides extra resources to disadvantaged preschool students. Head Start began in 1965 by giving \$1,000 (in 1999 prices) per student to about 500,00 children aged three and four; today it supports about

800,000 students with about \$5,400 per student (Garces et al. 2002). Analyses of the program show that Head Start increases test scores, lowers dropout and failure rates, and shrinks test score inequality between ethnic groups (Currie and Neidell 2003, Currie and Thomas 1999, Barnett 1995, and Karly et al. 1998). By third grade, however, the effects of Head Start seem to disappear (Currie and Thomas 2000, Aughinbaugh 2001; also see Barnett 1995 and Karly et al. 1998 for review of this literature). More recent analysis by Garces et al. (2002) has shown that adults who participated in Head Start as children were not more likely to complete high school, attend college, or have higher earnings than students who did not participate in Head Start. One suggested conclusion of this research is that in order to achieve long-term effectiveness, compensatory education must be sustained over long time periods (Aughinbaugh 2001).

Chile operates a program called P-900 that resembles CONAFE. Chile's Ministry of Education provides teacher training, textbooks and didactic materials, and infrastructure improvements to the schools with the worst performance on a national exam of student ability. Some recent analyses (MINEDUC 2000, Tokman 2002) compare performance of students in P-900 schools against students outside such schools. They find that P-900 increases test score performance over several years, but no analyses examine the effectiveness of P-900 in increasing school participation through the university level or in increasing wages through adulthood.

CONAFE's compensatory programs do not operate schools, but rather give extra support to all indigenous, some secondary, and selected other rural schools.³ For schools that enroll indigenous students, CONAFE supports development of curricula, didactic materials, and textbooks in an indigenous language and Spanish to facilitate bilingual education. CONAFE also

³ CONAFE does operate a community education program that leads instruction in highly isolated areas. Since we only examine CONAFE's compensatory programs, subsequent mention of "CONAFE" refers only to CONAFE's compensatory programs unless otherwise noted.

supports the development of intercultural education for indigenous students—CONAFE develops curricula that include elements from both Mexico’s mixed *mestizo* heritage and from indigenous cultures, and those curricula emphasize relationships between *mestizo* and indigenous culture. For disadvantaged rural schools, CONAFE provides updated audiovisual technology, professional development of teachers, improvements to school infrastructure, and other interventions designed to improve the learning outcomes of disadvantaged Mexican students. For *telesecundaria* education, CONAFE provides audiovisual materials and infrastructure improvements to all schools. In the *telesecundaria* schools that CONAFE supports, a single teacher coordinates and facilitates all subjects in a rural school that receives recorded lectures from Mexico City using a national satellite television system. In most CONAFE schools, a group of community parents and leaders (AGEs) receives a grant that can be spent on whatever educational purposes the group selects. The involvement of local decision-making may increase the effectiveness of spending; it may also increase parental involvement in and commitment to their students’ learning. For simplicity, we will describe schools and students that receive support from CONAFE as CONAFE schools and CONAFE students.

A targeting index selects which schools receive CONAFE support. Indigenous peoples in Mexico are sufficiently segregated and distant from even under-developed urban areas that separate schools exist for indigenous students. Most indigenous schools and *telesecundarias* receive CONAFE’s support. CONAFE selects other schools for support based on the average income of the school’s community, the school’s isolation and access to public infrastructure, the school’s education indicators and other indicators of poverty.

In the last 15 years, CONAFE has received substantial funding from international agencies. The World Bank’s Basic Education Loan (PAREIB, 1998-2006) provides a nominal

total of \$625 million to support CONAFE. The World Bank operated several similar loans between 1991 and 1998. The Inter-American Development Bank has operated the Integrated Program to Abate Educational Underachievement (PIARE) and the Distance Education Program. These loans provided a nominal total of nearly two billion dollars between 1991 and 2003 (Table 1).

Table 1. International Financial Support for CONAFE

Project name	Years	Amount	Details
First Primary Education Project (PARE)	1991-1997	\$250	World Bank. Supported compensatory programs for primary education in the four Mexican states with the highest incidence of poverty.
Initial Education Project (PRODEI)	1993-1997	\$74 ⁴	World Bank. Supported human resource development of educators; educational materials, and institutional strengthening
Integrated Compensatory Education Program (PIARE)	1995-2000	\$393	Inter-American Development Bank. Supported CONAFE's Compensatory Programs in the 17 states that PAREB did not cover.
Second Primary Education Project (PAREB)	1994-2001	\$412	World Bank. Supported compensatory education in 14 Mexican states.
Distance Education Program	1998-2003	\$171	Inter-American Development Bank. Supported <i>telesecundaria</i> education.
Basic Education Development (PAREIB)	1998-2006	\$625	World Bank. Supports basic compensatory education in all Mexican states.
Comprehensive Community Education Program	2003-	\$420	Inter-American Development Bank. Supported CONAFE's Community Education Programs in all Mexican states.

Amounts are in nominal US\$ million. Sources: World Bank 1991, World Bank 1994, Inter-American Development Bank 1997, World Bank 1997, World Bank 1998, World Bank 2002, Inter-American Development Bank 2003.

That international funding supports a relatively low-cost program. CONAFE's real costs have grown in the last decade, and CONAFE's compensatory programs now cost just over \$50 per student per year. Those costs compare to a typical cost of \$527 per *telesecundaria* student and \$477 per general middle school student (Castro, Wolff and García 2001) (See Table 2).

⁴ Originally \$80m, but in 1995, \$6m of the loan was canceled due to Peso devaluation.

Table 2. The Cost of CONAFE Students, 1994-2002

	1994	1995	1996	1997	1998	1999	2000	2001	2002
Real Cost	\$17	\$25	\$41	\$40	\$45	\$48	\$51	\$48	\$51
Annual Growth (%)		0.49	0.66	-0.02	0.11	0.07	0.06	-0.04	0.06

Note: Costs are per Mexican student, in 2002 U.S. Dollars, using an exchange rate of 0.103MXP = US\$1. Source: CONAFE data. Costs deflated by Banco de México's Consumer Price Index.

The World Bank and the Inter-American Development Bank (IDB) support CONAFE based on the argument that CONAFE improves the educational outcomes of Mexico's most disadvantaged students. A variety of previous evaluations have examined the effectiveness of CONAFE with some or no control group construction, and they have generally found that CONAFE improves short-term educational outcomes. A World Bank (2002) evaluation compared CONAFE-supported schools between the years 1992 and 1995 with comparable schools in the Mexican state of Michoacan that at the time received no CONAFE support. That evaluation found that CONAFE significantly increased test scores of indigenous and other rural students. CONAFE appeared to increase the test scores of indigenous students by 25 percent over the four-year period, and CONAFE appeared to increase rural non-indigenous test scores by 12 percent over the period. The evaluation concluded that full implementation of CONAFE could cause indigenous student performance increases of 45 percent to 90 percent, and rural school performance increases of 19 percent to 38 percent.

A separate World Bank (2002) evaluation summarized in the same document found that CONAFE's support significantly benefited the one million indigenous primary school students in CONAFE schools. The report found that indigenous students supported by CONAFE were catching up to their non-indigenous peers in test scores by about 10 percent per year. The evaluation did not control for relevant differences in background between indigenous and non-

indigenous students. An additional IDB-sponsored evaluation found that *telesecundaria* schools were effective in increasing math and language test scores. The report found that *telesecundarias* entirely eliminated the math performance gap between *telesecundaria* and non-*telesecundaria* students, though *telesecundaria* education was not as effective in Spanish instruction (Castro, Wolff and García 2001).

An additional evaluation by Paqueo and Lopez-Acevedo (2003) studied the effect of CONAFE support on rural and indigenous children using a production function, measuring performance by sixth grade Spanish test scores. The study found that CONAFE improved learning outcomes for poor students, but that it appeared to have no distinct effect for indigenous poor students.

Modeling Background and Achievement

Prior evaluations imperfectly measure CONAFE's effect because student socio-economic backgrounds differ markedly between CONAFE and non-CONAFE schools. For example, practically all indigenous students attend CONAFE schools, so one cannot determine the effect of CONAFE on indigenous students by looking at only indigenous students—one would have no appropriate group against which to compare those students. Construction of a control group of non-indigenous students, however, is difficult—indigenous students are more disadvantaged than their non-indigenous peers, and indigenous students may lack Spanish as a mother tongue. So we seek to develop a model that can use available data on student test scores, CONAFE support, and student background to distinguish the effect of CONAFE on student achievement from the effect of student background on student achievement.

We first consider the probability P that the school of a student i received CONAFE support in year k , representing CONAFE support as $c_{ik}=1$. That probability depends on a function f of the background characteristics X of the student and the student's community such as availability of public services, average literacy and average income:

$$P(c_{ik} = 1) = f(X_{ik}) \quad (1.0)$$

Although CONAFE supports a school and not a single student, we model CONAFE support as a function of both student and community background. Since CONAFE selects schools by the average disadvantage of the students attending that school and the community in which the school is located, equation (1.0) accurately models CONAFE's targeting.

Then for two students 1 and 2 of comparable backgrounds and similar communities, by equation (1.0), the probabilities that either student's school receives CONAFE support in year k are equal:

$$f(X_{1k}) = f(X_{2k}) \Rightarrow P(c_{1k} = 1) = P(c_{2k} = 1) \quad (1.1)$$

We further assume that the educational achievement θ of a student in the year k_1 is a function g of the background of both the student and student's community and the presence c of CONAFE support in the prior year k_0 :

$$\theta_{ik_1} = g(X_{ik_1}, c_{ik_0}) \quad (2.0)$$

In reality, school and teacher quality significantly affect student learning. School quality, however, is likely to closely correlate with student and community background. Since the best measure of school quality, θ , is the outcome we are trying to predict, we assume that the information in X captures information about school quality, and we include no school or teacher quality indicator in equation (2.0). We should emphasize that CONAFE support requires at least one time period to be effective, so c in time period k_0 affects θ in time k_1 .

So two students 1 and 2 with similar backgrounds in time period k_l and whose schools either both or neither received CONAFE support in time period k_0 should demonstrate the same educational achievement:

$$X_{1k_1} = X_{2k_1} \cup c_{1k_0} = c_{2k_0} \Rightarrow \theta_{1k_1} = \theta_{2k_1} \quad (2.1)$$

Similarly, if two students have similar backgrounds but one student's school receives CONAFE support and the other student's school does not, the effect Δ of CONAFE's support in time k_0 on student achievement in time k_1 equals the difference in achievement of the students in time k_1 .

$$X_{1k_1} = X_{2k_1} \cup (c_{1k_0} = 1 \cup c_{2k_0} = 0) \Rightarrow \theta_{1k_1} - \theta_{2k_1} = \Delta \quad (2.2)$$

Thus, to measure the effect of CONAFE's support on student achievement, we should compare students with similar background characteristics but where one student receives CONAFE support and the other does not. Although the students may show similar performance at first, we hypothesize that each year of CONAFE support should improve the performance of CONAFE students by Δ . In sum, CONAFE's effect in one year can be measured as the difference in score increases demonstrated by students of comparable backgrounds, only one of whom received CONAFE support in the prior year.

In reality, CONAFE's effect likely compounds or decreases over time, and a simple one-period model does not show CONAFE's true long-run effect. Our data, however, cover a span of five school years for primary school and three school years for secondary school, and we are unable to determine when most schools in the sample began receiving CONAFE support. Thus, we limit our model to a simple one-period comparison and emphasize the caveat that CONAFE's effects may in reality compound or decrease over time.

Methodology

To identify students with similar backgrounds, we use a propensity score matching algorithm that identifies comparable CONAFE and non-CONAFE students. Propensity score matching was first advanced by Rosenbaum and Rubin (1983) and is becoming more common in evaluations of targeted anti-poverty programs (see, for example, Pradhan and Rawlings 2002; Newman et al 2002; Park, Wang and Wu 2002; and Jalan and Ravallion 2003). Matching is particularly appropriate to evaluate cases where (a) some observations received an experimental treatment, (b) selection of observations for treatment was non-random and based on background features of each observation, (c) few observations in the non-experimental group have similar characteristics to observations in the experimental group, and (d) selecting comparable experimental and non-experimental observations is difficult due to the high number of background features needed to determine comparability (Dehejia and Wahba 2002). CONAFE's support is non-random and heavily based on complex background characteristics of students and schools, so evaluation of CONAFE is a good case for the matching evaluation.

An unbiased propensity score match would use sufficient background information of students to ensure that the assignment of CONAFE support among students with equal propensity scores is purely random (Becker and Ichino 2002). Unfortunately, limited background data available to do this match suggests that negative score bias towards CONAFE students remains in our methodology, meaning that CONAFE students must overcome extra obstacles to show the same achievement as non-CONAFE students.

A propensity score (p-score) is the probability, given a school's background, that the school receives CONAFE support. A p-score could more simply be interpreted as the correspondence of a school's background with the profile of a typical CONAFE school.

Our scoring methodology was based on Dehejia and Wahba (2002) and consists of five steps that the following pages explain in detail. First, we selected background variables to use in comparing students. Second, we used a probability function to estimate a p-score for each school, and assigned each student the p-score of his/her school. Third, we sorted students by their schools' p-score. Fourth, we separated students into groups so no significant difference existed between the mean p-score of the CONAFE and non-CONAFE students in a group. Finally, we tested separately for each background variable whether values of the variable were balanced between the CONAFE and non-CONAFE students in a group. We used this algorithm for primary and secondary schools separately.

Following this methodology, we first used three criteria to select background variables that would identify similar CONAFE and non-CONAFE schools in those databases. First, we excluded endogenous variables. For example, data on student failure rates would help identify similar CONAFE and non-CONAFE schools. But a goal of CONAFE is to lower failure rates. Since CONAFE likely affects the failure rates of its students, data on failure is an inappropriate control for isolating the effect of CONAFE. If all background data from years before when CONAFE began, then endogenous variables would effectively select CONAFE and non-CONAFE schools. Since we do not, however, have data on most CONAFE schools before CONAFE began supporting those schools, we exclude such variables from background selection.

Second, we required the availability of data on a variable for at least two different years and grades. Since a school's propensity score was constant for all five years, we wanted to avoid the potential bias of selecting an outlier data point from one grade-year.

The third criterion considered specific requirements of the propensity score match. To ensure balance between the CONAFE and non-CONAFE group, we had to exclude some

variables that perfectly predicted CONAFE participation and other variables that could not have equal sample means in p-score divisions of the CONAFE and non-CONAFE samples.

These three criteria gave us five context variables for primary schools and four context variables for secondary schools. For primary schools, we considered the Mexican state in which the school was located, the community's principal economic activity (primary, secondary or tertiary), the school modality (urban public, urban private, rural or indigenous), the average literacy of students' fathers, and the presence of sewage facilities. All background information came from student surveys except for information on community economic activity, which came from a school director's survey. For secondary schools, we considered the state of Mexico in which the school is located, the community's principal economic activity (primary, secondary or tertiary), the presence of a telephone in the student's community, and the average literacy of students' fathers. All indicators for both primary and secondary school were expressed as a series of dummy variables.

We then formed combined databases of schools for which we had data in any year. We considered including only schools for which data appeared in every year of the 1998-2002 sample. For the primary schools, that limit would have included about 70 percent of all the schools that appeared in any year. Because it is unlikely, however, that the schools surveyed in every year are randomly selected from the schools surveyed in any year, we instead formed a combined database of schools present in any year. We collapsed that database to include one entry for each school, including mean values for the selected background variables. That collapse produced one database of schools with codes and background information for each school.

With the school database, we then used a probit model to determine the propensity score (p-score) for each school. The probit estimates the probability that a school will receive CONAFE support given a vector of the selected context variables:

$$P(CONAFE_{i,t} = 1) = F(Q_{i,t}) \quad (3.1)$$

$$P(CONAFE_{i,t} = 0) = 1 - F(Q_{i,t}) \quad (3.2)$$

$$Q_{i,t} = \phi_t^0 + \phi_t^1 Z_{i,t}^P + \varepsilon_{i,t} \quad (3.3)$$

where CONAFE is a dummy variable indicating whether school i in year t received CONAFE support; and $Q_{i,t}$ is an index function based on a vector of school background characteristics $Z_{i,t}$, an error term ε , and a series of constants ϕ . This probit gave us the p-score of each school observation.

At this point we associated students with the p-scores of their schools, so we had a database of students and each student had a p-score. We then ranked student observations by ascending p-scores and divided observations into groups so the difference in mean p-score between CONAFE and non-CONAFE schools for a given group was not significantly different at the 0.05 level. One group, for example, included all students with p-score between 0.0 and 1.1. We then tested that within each group, the difference in mean value for each background variable was not significantly different between CONAFE and non-CONAFE schools. When we could not satisfy that balancing property, as explained in the third criterion of variable selection, we increased the number of groups; when such increase could not satisfy the balancing property, we excluded the variable from construction of p-scores.

Since not every disadvantaged school receives CONAFE support, the p-score effectively selects CONAFE and non-CONAFE schools with similar backgrounds. A p-score of 1 indicated that a school had a typical CONAFE profile (i.e., student's parents are illiterate, school is small

and rural, etc.), and a p-score of 0 indicated that the school had a background very different from the background of a typical CONAFE school.

It is interesting to know not only how CONAFE affects student achievement, but to know how CONAFE differently affects the most and least disadvantaged students in Mexico. To make such comparison, we divided students into groups of relative disadvantage with p-scores between 0.1 and 0.3, p-scores between 0.4 and 0.6, and for primary schools, p-scores between 0.7 and 0.9, and we compared performance over time of matched CONAFE and non-CONAFE students by relative disadvantage. For example, we compared the improvement of matched CONAFE and non-CONAFE students with levels of disadvantage between 0.4 and 0.6 between 1998 and 2002. According to equation (2.2), the amount by which CONAFE student improvement exceeded non-CONAFE student improvement represented the effect of CONAFE on student learning.

Data

We used three types of databases for this evaluation: student-level test databases (Test) containing test scores and student background information, school-level databases (Drop) containing data on student repetition and failure, and school-level background databases (Bgrd) with data from school principal and school teacher questionnaires. Entries in each of the three databases included a school code that let us identify the same school in all three databases. A summary of databases used appears in Tables 3 and 4 below.

Table 3. Primary School Databases

Respondent	Year(s)	Type	Grade	Obs	Schools
Principal	1996-2000	Bgrd		3,681	3,681
Student	1998	Test	Second	43,398	2,581
Student	1998	Bgrd	Fourth	47,497	3,140
Student	1998	Bgrd	Fifth	44,912	3,037
Student	1999	Test	Third	49,136	3,098
Student	2000	Test	Fourth	46,252	3,040

Student	2000	Bgrd	Sixth	45,676	3,189
Student	2001	Bgrd	Third	48,903	3,268
Student	2001	Bgrd & Test	Fifth	46,718	3,221
Student	2002	Bgrd	Fourth	48,451	3,205
Student	2002	Bgrd & Test	Sixth	46,365	3,165
Student	1992-2002	Drop	First-Sixth	583,876	82,520
Teacher	1998	Bgrd		23,357	3,562
Teacher	2001	Bgrd		22,917	3,437
Teacher	2002	Bgrd		20,634	3,373

Source: SEP data.

Table 4. Secondary School Databases

Respondent	Year	Type	Grade	Obs	Schools
Principal	2000	Bgrd		1,406	1,403
Principal	2001	Bgrd		1,184	1,181
Principal	2002	Bgrd		860	860
Student	2000	Test	Seventh	39,562	1,189
Student	2001	Bgrd & Test	Eighth	39,562	1,189
Student	2002	Bgrd	Eighth	53,919	1,982
Student	2002	Bgrd & Test	Ninth	53,202	1,980

Source: SEP data.

Information on student test scores came from *Estándares Nacionales* (EN), an exam that Mexican students took in each year during the period 1998-2002. This test provides information about student ability in Spanish and math. Mexico's Secretariat of Public Education (SEP) also reports a composite global score based on the student's Spanish and math ability. SEP standardizes scores to control for changes in exam difficulty between grades and test years. EN gives information on 45,000 students from 3,000 primary schools. The EN sample is relatively small: in 1999, CONAFE gave didactic content support to about 4 million students, so the EN sample includes about 1 percent of all CONAFE students.

EN's sampling is based on an earlier selection of representative schools. In 1996, SEP selected 3,000 schools to include in its test-score sample.⁵ Those schools evenly represented states, socioeconomic backgrounds, student gender, and other demographic factors of the

⁵ In the 1996-1998 period, SEP gave students a different test, the results of which are not comparable with EN.

Mexican population. Each year, SEP adjusts the schools and students in the sample to ensure that it represents Mexico's national profile.

For choosing students within the selected schools, SEP uses two algorithms. For primary schools with fewer than 80 students, SEP included every student in the sample who took the exam. Since SEP gave exams in few grades each year, not every student in the school appeared in the exam. For primary schools with more than 80 students, SEP randomly selected 25 students to include in the sample. For secondary schools with fewer than 200 students, SEP included every student in the sample who took the exam. For secondary schools with more than 200 students, SEP randomly selected 36 students to include in the sample. So the likelihood that a student in a small school appears in the sample is higher than the likelihood that a student in a large school appears in the sample. This bias suggests that our results over-represent students in small schools.

The EN database included data from a variety of grades and years. To allow comparison of a cohort, we used primary school test data from second grade in 1998, third grade in 1999, fourth grade in 2000, fifth grade in 2001 and sixth grade in 2002. Those grades represent the grade of test content; a fourth grade exam includes material that a student should have learned in fourth grade. Similarly, we used secondary school test data from third grade in the year 2000 through third grade in the year 2002. The exact students in each grade-year varied, so we did not exactly use a cohort. Rather, we followed a group of schools through time while allowing the students within those schools to vary randomly.

For each student in the EN sample, the Evaluation Division of Mexico's Secretariat of Public Education (DGE) reports three scores. The first is a Spanish score based on a multiple-choice Spanish language exam. The second is a mathematics score also based on a multiple-

choice exam. The third is a global score, which is a composite number based on Spanish and math scores that represents a student's overall exam performance. While an increase in math or Spanish scores will increase global scores, the increase is rarely proportional.

SEP calibrates exams using the Rasch correction, which compensates for changes in exam difficulty in different years and grades. In the Rasch model, the probability of a student i giving a correct response to question j , represented as $P(y_{ij}=1)$, depends on the student's ability ϕ_i and the difficulty of the question θ_j . So the probability of a correct response to question j is the following:

$$P(y_{ij} = 1) = \frac{e^{[(\phi_i - \theta_j)y_{ij}]} }{1 + e^{(\phi_i - \theta_j)}} \quad (3.1)$$

That correction compensates only for changes in the difficulty of exams in different years and grades. Thus, a score of 150 in Spanish in 1998 represents the same knowledge level as a 150 in Spanish in 2000, even though the exam questions in 2000 may have been more difficult than the questions in 1998. If student ability improves over time, then mean test scores should rise over time. The correction also makes different grades comparable: a student's 150 in Spanish in third grade one year and her 150 in Spanish in fourth grade the next year suggest that the student did not learn in fourth grade. Since this correction does not compensate for changes in student learning, average scores should increase over time and between grades.

The Drop databases list failure and repetition rates by year for every school that appears in the Test database. While the Test data are at the student level, the Drop data are at the school level. Also, the Drop data are averaged for all grades in the school, so Drop lists only a single repetition and dropout number for each school.

The third data source includes a variety of databases that include context data from questionnaires given to teachers and school principals (Tables 3-4). While we can identify a

student, parent and school principal in different databases with a particular school, we cannot identify individual observations in different databases with each other.

Sample bias

Between 2 and 12 percent of students who took the exams did not provide some context data. Those non-responding students were significantly more disadvantaged than were responding students. In a given year, the portion of CONAFE students who failed to provide information was between 1.5 and 3 times the portion of non-CONAFE students who failed to provide information. Since much context information was available at the school and not the student level, we had some context data for most students.

We did simple explorations of non-response bias by comparing the responses for one background indicator between students who did and did not respond to another indicator. For example, we compared the failure rates of students who did and did not indicate whether their parents were literate (Table 5). It is unclear whether non-response is due to students purposefully not filling out bubbles on answer sheets, administrative errors in processing student replies, or other reasons. We made this comparison for several grades and indicators; only two grades appear in Table 5. In every grade, the rate of non-response among CONAFE students was 1.5 to 3 times the rate of non-response among non-CONAFE students. The students who did not reply to the literacy question had markedly higher rates of failure than students who did reply to the failure question. Failure rates among students who did not reply to the literacy question were 20-50% higher than were failure rates of students who did reply to the literacy question.

Table 5. Failure rates among students who gave no literacy reply

<i>Percent</i>	<i>1998 Fourth Grade</i>			<i>2000 Sixth Grade</i>		
	<i>Con</i>	<i>Non</i>	<i>Diff</i>	<i>Con</i>	<i>Non</i>	<i>Diff</i>
Gave no failure reply	12.02	4.97	7.05	4.08	2.51	1.57
Overall failure rate	0.33	0.18	0.15	0.32	0.17	0.15
Failure rate if gave no literacy reply	0.39	0.29	0.09	0.43	0.29	0.15
Difference	0.06	0.12	0.06	0.12	0.12	0.00

Source: SEP Data

We performed similar comparison for the literacy rates of students who gave no failure reply, and for the global scores of students who gave no failure reply. In every case, the rate of non-response among CONAFE students was significantly higher than the rate of non-response among non-CONAFE students. It also appeared that the bias created by non-response for a given student among CONAFE and non-CONAFE students was comparable. In other words, the difference between responding and non-responding students was comparable among both CONAFE and non-CONAFE students. Since a higher portion of CONAFE students failed to provide background information, and since non-responding students were more disadvantaged, we interpreted that non-response bias made the CONAFE students more disadvantaged than the non-CONAFE students. For CONAFE results in the present paper to equal non-CONAFE results, then, CONAFE students must overcome the additional disadvantage of their backgrounds that response bias prevents us from measuring. This finding further highlights the caveat that results of this report are an absolute lower bound for the positive impact of CONAFE on student achievement.

Descriptive Statistics

In a given grade-year, CONAFE supported 35-45 percent of primary school students in the sample; and 30-40 percent of secondary schools a given grade-year were *telesecundarias*. CONAFE support was also relatively static over the sample period. Nearly 70 percent of schools

in the sample received CONAFE support for all five years of the sample. Ten percent of schools received CONAFE support for four years, 5 percent received support for three years, 10 percent received support for two years, and 7 percent received support for only one year. CONAFE also appeared to give the longest-standing support to the most disadvantaged schools, as parent literacy, community services and other disadvantage indicators showed more extreme levels of disadvantage in schools that CONAFE supported for longer periods of time.

Though CONAFE's compensatory education programs operate in every state, the number of CONAFE-supported schools varies markedly between states. For example, over 20 percent of the CONAFE schools in the sample come from just four states – Chihuahua, Veracruz, Jalisco and Durango – while the six states of Baja California, Mexico City, Guanajuato, Tlaxcala, Morelos and Zacatecas combined include just 10.7 percent of CONAFE schools. These data are based on the number of CONAFE schools in a state relative to the number of CONAFE schools in all of Mexico, so the number of schools in a state may affect a state's ranking.

CONAFE schools are significantly poorer and more disadvantaged than non-CONAFE schools are. In 1998, for example, 93 percent of non-CONAFE mothers were literate, whereas 78 percent of CONAFE mothers were literate. In the same year, only 39 percent of non-CONAFE community employment centered on primary (extractive) industries, while 89 percent of CONAFE community employment centered on primary industries. Measuring CONAFE's effect by comparing CONAFE scores to non-CONAFE scores assumes that students in the CONAFE and non-CONAFE group have similar backgrounds. CONAFE students live in less educated communities and have access to fewer social services than non-CONAFE students.

Primary School Results

Before presenting results, we should emphasize that the limited data available to pair students of similar backgrounds makes it likely that unobserved factors for which available data did not control could likely influence the results of this control group comparison. CONAFE students are generally more disadvantaged than are non-CONAFE students, so it is likely that the non-CONAFE students in each of the three disadvantage groups have more prosperous backgrounds than do CONAFE students. For the experimental and control groups to demonstrate comparable performance, CONAFE would have to overcome these unobserved factors. Since unobserved factors limit CONAFE results, figures based on control group analysis are an absolute lower bound for the positive effect of CONAFE on student achievement. In other words, the actual effect of CONAFE on student performance is likely more positive than this report estimates.

Global, math and Spanish scores of CONAFE students increased over the sample period for all three groups of disadvantage. While average global scores of the less disadvantaged and mid-range groups increased by 18.7 and 20.4 points per year, respectively, global scores of disadvantaged CONAFE students increased by only 16.3 points per year (Table 6). Disadvantaged students may learn more slowly because they have less pre-school education, which could cause their scores to increase at a slower rate. Of course, these increases are predictable given that the cohort includes higher grades in later years, and thus one should expect test scores to rise each year. The most accurate conclusion to infer from annual increase in test scores is that students gain new Spanish and math skills in each year of education, and that both CONAFE and non-CONAFE schools add value. We cannot, however, determine from this finding that CONAFE itself is a source of that value.

We also determined that CONAFE had caused part of these test score increases. In each of the three disadvantage groups, the experimental group gained annually on the control group by 2.4 to 4.3 points in global scores. This increase is the minimum effect attributable to CONAFE. Similarity in results across each of the three disadvantage groups re-affirms the positive effect of CONAFE on student achievement. CONAFE students gained on their non-CONAFE peers in math scores by 4.8 to 5.6 points per year. According to this data, however, disadvantaged CONAFE students gained on their peers by 1.2 points per year in Spanish exams, but mid-range and less-disadvantaged students fell behind by 2.8 to 3.6 points per year in Spanish exams (Table 6). These numbers may be negative due to the imperfect control group construction—unobserved factors correlated with CONAFE mean that these figures are the minimum positive effect attributable to CONAFE. They may also be negative simply because CONAFE did not increase test scores of all students. So CONAFE improved math scores in primary school, but its effect on Spanish scores was unclear.

One of main our findings is that CONAFE also appeared to decrease test score inequality between comparable CONAFE and non-CONAFE students. Disadvantaged non-CONAFE students outperformed their CONAFE peers on an average exam application by 13.7 global points. But CONAFE decreased global test score inequality between disadvantaged CONAFE students and disadvantaged non-CONAFE students by 4.2 points per year. A linear projection of CONAFE and non-CONAFE improvements shows that CONAFE decreased test score inequality in each category of exam and group of disadvantage by 9 percent per year in the less-disadvantaged group, 16 percent per year in the mid-range group, and 30 percent per year in the disadvantaged group (Table 6). Of course, actual test score improvements may be logarithmic, and it is likely that the significant test score increases demonstrated in the early years of

CONAFE support will shrink as the achievement of CONAFE students approaches the achievement of their non-CONAFE peers. Nonetheless, it seems that CONAFE is rapidly decreasing learning inequality, as CONAFE significantly decreases achievement inequality, and that the decrease in inequality is largest for the most disadvantaged students.

In all three disadvantage groups, CONAFE was slower to close the Spanish test score gap than to close the math test score gap between CONAFE and non-CONAFE students. In the mid-range and less-disadvantaged groups, this data showed no evidence of CONAFE closing the Spanish gap. While the Spanish gap of the disadvantaged group decreased at an annual rate of 1.2 points, the Spanish gap of the mid-range and less-disadvantaged groups increased at 2.7 and 3.3 points per year (Table 6).

One might hypothesize that CONAFE is less effective in primary school Spanish instruction than in math instruction due to the presence of indigenous students. This interpretation is incorrect. On a disadvantage ranking of 0 to 1, we defined the disadvantaged group to have a score between 0.7 and 0.9. Only 1 of the 2,744 indigenous schools in DGE's sample fell into the disadvantaged group. All other indigenous schools had a disadvantage ranking above 0.9. It is possible that indigenous students are present in schools identified as rural public schools. Those students would have a lower return to Spanish education. Hence, one would expect that CONAFE schools containing indigenous students would show slower improvement on Spanish exams than would CONAFE schools without indigenous students. But the Spanish-math differential is notable in both the disadvantaged and less-disadvantaged groups, and indigenous students are likely to have little presence in the less-disadvantaged group. Furthermore, Spanish education appeared to be more effective in disadvantaged schools than in less-disadvantaged schools, and indigenous students are likely to have the highest presence in

disadvantaged schools. Thus, influential factors besides the presence of indigenous students cause CONAFE to be more effective for math instruction than for Spanish instruction. Those factors may be pedagogical: perhaps improving math learning at the primary level requires few resources, while improving Spanish learning requires many resources. Also, the average gap between CONAFE and non-CONAFE students was larger in math than in Spanish for two of the three disadvantage groups studied. Thus, another possible explanation is simply that math instruction in non-CONAFE schools is less effective than Spanish instruction in non-CONAFE schools, so a given infusion of CONAFE resources will increase math more than Spanish scores.

Table 6. Primary School Test Scores, 1998-2002

	1998	1999	2000	2001	2002	Ave	Trend
<u>Less-disadvantaged group</u>							
Global							
Non-Conafe	452.8	431.4	481.7	489.3	505.2	472.1	16.3
CONAFE	414.9	410.2	456.5	463.6	481.7	445.4	18.7
Difference	-37.9	-21.3	-25.2	-25.7	-23.5	-26.7	2.4
Math							
Non-Conafe	453.6	431.8	483.8	488.4	498.0	471.1	14.5
CONAFE	416.2	405.5	453.4	466.7	483.0	445.0	19.5
Difference	-37.5	-26.3	-30.4	-21.6	-15.0	-26.2	5.0
Spanish							
Non-Conafe	416.3	429.5	479.9	490.6	510.9	465.4	25.0
CONAFE	396.8	412.9	460.0	457.9	481.3	441.8	21.4
Difference	-19.5	-16.6	-19.9	-32.6	-29.6	-23.6	-3.6
<u>Mid-range Group</u>							
Global							
Non-Conafe	451.0	428.9	481.4	488.8	501.7	470.3	16.1
CONAFE	408.3	412.2	450.3	463.3	485.1	443.8	20.5
Difference	-42.6	-16.7	-31.1	-25.5	-16.6	-26.5	4.3
Math							
Non-Conafe	449.5	428.8	482.5	488.9	495.8	469.1	15.3
CONAFE	408.8	410.6	450.3	464.7	486.0	444.1	20.9
Difference	-40.8	-18.2	-32.1	-24.2	-9.8	-25.0	5.6
Spanish							
Non-Conafe	410.3	426.9	480.7	488.4	506.4	462.5	25.4
CONAFE	395.5	411.7	450.0	459.4	484.7	440.3	22.6
Difference	-14.8	-15.2	-30.7	-29.1	-21.6	-22.3	-2.8
<u>Disadvantaged Group</u>							
Global							
Non-Conafe	458.2	423.2	467.1	478.7	491.2	463.7	12.1
CONAFE	430.8	410.2	457.7	467.2	483.9	450.0	16.3
Difference	-27.4	-12.9	-9.4	-11.5	-7.3	-13.7	4.2
Math							
Non-Conafe	458.0	420.6	466.3	478.2	488.4	462.3	11.8
CONAFE	434.2	404.6	457.2	469.8	484.9	450.1	16.7
Difference	-23.8	-16.0	-9.1	-8.5	-3.5	-12.2	4.8
Spanish							
Non-Conafe	416.1	424.0	467.4	478.3	493.4	455.9	20.9
CONAFE	397.2	413.5	457.9	462.3	483.2	442.8	22.1
Difference	-18.9	-10.6	-9.5	-16.1	-10.3	-13.1	1.2

Source: *Estándares Nacionales 1998-2002*

Indigenous Primary School Results

Since indigenous students tend to be geographically isolated, SEP identifies indigenous education as a modality separate from rural or urban education. But the difficulty of forming an effective control group against which to compare indigenous students makes evaluation of indigenous education difficult. Every indigenous school but one had a p-score of over 0.9. Those scores are high in part because nearly all indigenous schools receive CONAFE support, and in part because indigenous schools are extremely disadvantaged. Because the schools with p-scores over 0.99 represent the highest tail end of disadvantage, it is likely that indigenous schools in this group have very different characteristics than do non-indigenous schools in this group. Thus, this section focuses on the indigenous school group with p-scores between 0.9 and 0.99. We use non-indigenous and non-CONAFE schools with p-scores between 0.9 and 0.99 as control groups. Data from 1998 in the 0.9 to 0.99 p-score range included only 55 indigenous students, so we excluded 1998 data from this analysis.

Indigenous students effectively learned much of the material they studied in primary school. Indigenous students with CONAFE support rapidly improved their test performance over the sample period. Indigenous global scores increased an average of 19 points per year, with an annual average math score increase of 26.8 points and an average Spanish score increase of 12.3 points. From 1999 to 2002, indigenous scores increased from 398 to 454 points in global scores, from 385 to 464 points in math scores, and from 407 to 447 points in Spanish scores (Table 7). Again, it is likely that much of this increase is attributable to cohort's following of students through different grades in different years, so the appropriate conclusion to reach from this data is that indigenous students learn a significant amount of new material in each year of education.

CONAFE appeared to cause some of this math score improvement, but again we found no effect on Spanish scores. In 1999, indigenous students scored an average of 32.6 points behind comparable non-CONAFE non-indigenous students did in math scores. That math score gap decreased by an average of 5 points annually. Due to yearly test score fluctuations, however, indigenous students in 2002 remained 21.2 points behind comparable non-indigenous students. Decomposing non-indigenous students into CONAFE and non-CONAFE students gives a clearer picture. Indigenous students remained significantly behind their non-indigenous CONAFE and non-CONAFE peers. In math scores, however, indigenous students gained on non-indigenous CONAFE students by a 5.0 points per year, and they gained on non-indigenous non-CONAFE students by 6.5 points per year (Table 7).

Indigenous students actually fell behind their peers in Spanish scores, as indigenous student Spanish scores increased from an average 407 points in 1999 to 447 points in 2002, with an average annual increase of 12.3 points per year. But on the same exam over the same time period, non-indigenous students improved by an average annual increase of 19.1 points per year from 419 to 481 points (non-indigenous CONAFE) and from 429 to 493 points (non-indigenous non-CONAFE) (Table 7). At least two factors explain why indigenous Spanish learning improves slowly. First, many indigenous students have a native language as mother tongue and do not receive reinforcement of school Spanish learning at home. Since practically all indigenous students receive CONAFE support, available data do not allow us to control for home language in measuring CONAFE's effect. Second, instruction of indigenous students faces extra hurdles of the difficulty of finding well-trained bilingual teachers and bilingual didactic materials. CONAFE has focused on overcoming the second obstacle to indigenous student performance, and results show that indigenous students are improving their Spanish performance. But because

practically every indigenous student is in CONAFE, the most accurate conclusions one can reach are that indigenous student performance on Spanish and math exams is improving, that indigenous students are learning more quickly in math than in Spanish, and that indigenous students are catching up to their peers in math.

Table 7. Primary School Test Scores by Ethnicity, 1999-2002

	1999	2000	2001	2002	Ave	Trend
Global						
Indigenous	397.5	415.0	436.4	453.6	425.6	19.0
<u>All non-ind</u>	417.8	461.8	464.8	484.2	457.2	20.2
Non-CONAFE	425.5	468.8	467.3	490.7	463.1	19.4
CONAFE	417.2	460.0	464.3	482.1	455.9	19.9
Diff	20.3	46.8	28.4	30.7	31.5	1.3
Math						
Indigenous	385.4	407.0	440.4	463.6	424.1	26.8
<u>All non-ind</u>	414.5	460.2	467.4	484.8	456.7	21.8
Non-CONAFE	421.4	465.0	470.2	487.2	460.9	20.3
CONAFE	414.1	459.0	466.7	484.0	456.0	21.8
Diff	29.2	53.2	27.0	21.2	32.6	-5.0
Spanish						
Indigenous	407.3	424.7	428.0	447.3	426.8	12.3
<u>All Non-indig.</u>	419.3	463.1	459.6	484.0	456.5	19.1
Non-CONAFE	428.5	472.8	462.1	493.4	464.2	18.4
CONAFE	418.7	460.8	459.1	480.9	454.9	18.5
Diff	12.0	38.5	31.7	36.7	29.7	6.7

Source: *Estándares Nacionales* 1999-2002. "Diff" is the difference between indigenous and all non-indigenous students. Universe is students with p-score between 0.9 and 0.99.

Secondary School Results

Mexico offers both traditional education and *telesecundaria* education for secondary students. *Telesecundarias* are rural schools where lectures are played on television for students and a local teacher oversees classwork and exam application. This section only examines *telesecundaria* schools.

Math and Spanish scores of *telesecundaria* students increased over the sample period. Global scores of both disadvantage groups rapidly increased from an average of 499 points in 2000 to an average of 540 points in 2002. A linear projection of that trend showed that global scores of the less-disadvantaged *telesecundaria* group increased by 21.5 points annually, and that scores of the mid-range group increased by 19.3 points annually. Similar trends appeared for math and Spanish exams. For the less-disadvantaged group, math scores rose by 10.7 points annually and Spanish scores rose by a striking 27.2 points annually. Similarly, for the mid-range group, math scores rose by 12.7 points annually and Spanish scores rose by 26.7 points annually (Table 8). These increases are partly attributable to the different grades of students in this cohort in different years.

Telesecundaria education appeared to cause part of those increases in math and Spanish scores, as *telesecundaria* students showed more rapid score improvement than comparable non-*telesecundaria* students did in math and Spanish exams during the 2000-2002 period. For the less-disadvantaged group, *telesecundaria* average annual performance growth exceeded non-*telesecundaria* average annual performance growth by 1.4 points in math and by 3.4 points in Spanish. Results were less clear in the mid-range group, where inequality between *telesecundaria* and non-*telesecundaria* students decreased by 0.02 points per year in math scores and 0.8 points per year in Spanish scores. *Telesecundaria* education eliminated 24 percent of

average *telesecundaria*-non-*telesecundaria* inequality in math scores, and 33 percent of average inequality in Spanish scores. The effects of *telesecundaria* education were less notable in the mid-range group, where *telesecundarias* eliminated 0.4 percent of average inequality in math scores and 10 percent of inequality in Spanish scores (Table 8). These numbers are the minimum positive effect of *telesecundarias* on improving student achievement.

For both groups of disadvantage, *telesecundarias* appeared to be more effective in improving Spanish scores than in improving math scores. This finding is particularly interesting given that CONAFE's primary school interventions appeared to be more effective for math education than for Spanish education. Data on secondary education lacks a modality identifier, so we cannot determine the portion of indigenous students in this sample. Since no high-disadvantage group could be constructed, however, it is likely that few or no indigenous schools appear in this sample, and that other factors must explain the difference in performance between Spanish and math exams. This finding contrasts with an IDB evaluation, which found that *telesecundaria* education was effective in both math and Spanish instruction, but that *telesecundarias* were slightly more effective in math than in Spanish (Castro, Wolff and García 2001).

It may be that self-selection bias of rural students biases analyses of *telesecundarias*—perhaps the best-achieving rural students make extra effort to attend *telesecundarias*, while lower-achieving rural students do not. If those students' observable background characteristics reflect their propensity to learn in school, then the p-score matching should eliminate this potential bias. If observable characteristics do not identify student propensity to learn, however, then this bias may affect these results, and imply that the CONAFE students discussed here have

a higher propensity to achieve than their backgrounds imply, and thus CONAFE may not explain all of the *telesecundaria* score improvement.

Table 8. Test Scores of *Telesecundaria* students

	2000	2001	2002	Ave	Trend
<u>Less-Disadvantaged Group</u>					
Global					
Non-tel	530.2	552.5	564.2	549.0	17.0
Tel.	523.9	539.2	562.6	541.9	19.3
Difference	-6.3	-13.3	-1.6	-7.1	2.4
Math					
Non-tel	527.8	534.2	546.4	536.1	9.3
Tel.	521.9	525.7	543.3	530.3	10.7
Difference	-5.9	-8.5	-3.1	-5.8	1.4
Spanish					
Non-tel	535.9	570.4	583.6	563.3	23.9
Tel.	528.1	552.6	582.5	554.4	27.2
Difference	-7.8	-17.8	-1.1	-8.9	3.4
<u>Mid-Range Group</u>					
Global					
Non-tel	481.19	495.67	522.66	499.84	20.73
Tel.	473.98	488.58	517.05	493.2	21.53
Difference	-7.2	-7.09	-5.61	-6.64	0.8
Math					
Non-tel	497.67	507.98	523.08	509.58	12.71
Tel.	492.4	503.18	517.86	504.48	12.73
Difference	-5.27	-4.8	-5.23	-5.1	0.02
Spanish					
Non-tel	473.34	495.02	525.14	497.83	25.9
Tel.	465.27	486.94	518.65	490.28	26.69
Difference	-8.07	-8.08	-6.49	-7.54	0.79

Source : *Estándares Nacionales 2000-2002*.

Failure and Repetition Rates

A main goal of CONAFE is to lower failure and repetition rates. CONAFE maintains databases on these indicators for all schools, which allowed us to measure CONAFE's effect on failure and repetition rates.

We began with data on repetition and failure rates in every school that appeared in DGE's sample. We compared repetition rates for a cohort of CONAFE and non-CONAFE schools by disadvantage group for second grade in 1998 through sixth grade in 2002. The

average annual change in inequality between the CONAFE and non-CONAFE groups is the minimum positive effect attributable to CONAFE.

Repetition rates of CONAFE students were, as expected, significantly higher than repetition rates of non-CONAFE students in every grade, year, and disadvantage group. Repetition rates dropped by between 0.42 and 0.66 percentage points per year among CONAFE students (Table 9). Of course, because our analysis compares different grades in different years, it is likely that the lowest achieving students leave school each year, and that a decrease in repetition rates is partly attributable to difference in the students that attend each grade.

But it appears that CONAFE caused part of this decrease in repetition rates. In the disadvantaged group, CONAFE students had an annual average increase of 0.08 percentage points in repetition rates above their non-CONAFE peers. But in both the mid-range and non-disadvantaged groups, CONAFE students had an annual average decrease of 0.35 and 0.13 percentage points, respectively, against their non-CONAFE peers. Those decreases represent the lower level of repetition that CONAFE causes among its supported students. On balance, it appears that CONAFE decreases repetition rates by about 0.13 percent (0.0013 points) per year, but that effect varies between disadvantage groups. While this effect may appear to be small, it suggests that CONAFE annually eliminates 6 percent of inequality in repetition rates between comparable CONAFE and non-CONAFE students (Table 9).

Data on failure rates suggest less conclusive results. CONAFE students fail at a much higher rate than do comparable non-CONAFE students: among the grades and disadvantage groups analyzed, an average of 5.04 percent of non-CONAFE students failed a class subject in a given year, whereas an average of 7.22 percent of CONAFE students did. But the annual effect of CONAFE on decreasing failure rates appeared to be positive for the non-disadvantaged group,

negative for the mid-range group, and nonexistent for the disadvantaged group. The average of those groups indicates that CONAFE lowers failure rates annually by 0.3 percent (.003 points), but that number is negative primarily because the non-disadvantaged group showed such a significant effect on failure rates (Table 9).

Table 9. Effect of CONAFE on Failure and Repetition Rates, 1998-2001

	1998	1999	2000	2001	Ave	Trend
Less-disadvantaged Group						
Failure						
Noncon	4.38	4.03	3.88	3.47	3.94	-0.29
Con	9.65	7.12	6.28	5.36	7.1	-1.37
Diff	5.27	3.09	2.4	1.9	3.16	-1.08
Repetition						
Noncon	4.42	3.85	3.85	3.47	3.9	-0.29
Con	7.66	7.52	6.45	6.62	7.06	-0.42
Diff	3.25	3.66	2.59	3.16	3.16	-0.13
Mid-range Group						
Failure						
Noncon	6.58	6.21	5.3	5.09	5.79	-0.54
Con	7.61	7.25	6.94	6.54	7.08	-0.35
Diff	1.03	1.04	1.64	1.45	1.29	0.19
Repetition						
Noncon	5.3	5.67	5.44	4.34	5.19	-0.31
Con	8.44	6.66	7.16	6.08	7.08	-0.66
Diff	3.14	0.98	1.72	1.74	1.89	-0.35
Disadvantaged Group						
Failure						
Noncon	6.66	4.55	5.37	5.02	5.4	-0.41
Con	7.72	8.25	7.26	6.66	7.47	-0.42
Diff	1.06	3.71	1.89	1.65	2.08	0
Repetition						
Noncon	7.39	4.89	6.22	5.09	5.9	-0.56
Con	7.87	7.59	7.28	6.36	7.28	-0.48
Diff	0.48	2.7	1.06	1.27	1.38	0.08

Source: *Estándares Nacionales* 1998-2001. Data are percentages of all students in sample. "Failure" is the percent of students that have failed a grade. "Repetition" is the percent of students that have failed a subject or grade.

Conclusions

This analysis used student test scores and school-level indicators to evaluate the effectiveness of CONAFE in improving the quality of primary and *telesecundaria* education. We have used five approaches to measure CONAFE's effects. First, we compared the disadvantage of all CONAFE and non-CONAFE schools to examine the effectiveness of CONAFE's targeting

mechanism and to show the need for construction of control groups. Second, we measured CONAFE's effect on primary school test scores. Third, we measured CONAFE's effect on indigenous students. Fourth, we measured CONAFE's effect on secondary school test scores. Finally, we analyzed CONAFE's effect on failure and repetition rates.

CONAFE's targeting mechanism appears to be accurate. CONAFE's guidelines direct that every indigenous school should receive CONAFE support. We found that only 18 schools in the sample, representing 0.5 percent of indigenous schools, did not receive CONAFE support, and those schools may simply be data errors. CONAFE students are significantly more disadvantaged than are non-CONAFE students in test score achievement, parent literacy, household services and community industrial development. These differences between CONAFE and non-CONAFE students show that comparison of CONAFE performance to national performance fails to control for important background differences, and that construction of a comparable non-CONAFE control group is necessary to measure CONAFE's effect.

In primary schools, we found that CONAFE students effectively learned new material throughout primary school and that CONAFE added value in math instruction, but we found no significant impact of CONAFE on Spanish scores. CONAFE increased math scores by 4.8 to 5.6 points per year, and CONAFE decreased inequality between comparable CONAFE and non-CONAFE students in global scores by between 9 and 30 percent per year. CONAFE's insignificant effect on primary school Spanish learning is not due to the presence of indigenous students in CONAFE schools. Also, our methodology of control group construction meant that these numbers are the minimum positive effect attributable to CONAFE, and so CONAFE's true effect is above these estimates. Overall, CONAFE caused an annual score increase of 3.7 global points, or 0.8 percent. Given average global score inequality between CONAFE and non-

CONAFE students of 22.3 points, a linear projection of score increases would predict CONAFE to eliminate test score inequality in about six years. It is likely, however, that test score increases are logarithmic and not linear, and thus actual time required for test scores to become equal may be much longer.

Indigenous students learn a significant amount of new material as they proceed through primary school. CONAFE students gained in math scores by 6.5 points against non-CONAFE non-indigenous students and by 5 points annually against comparable CONAFE non-indigenous students. However, CONAFE had no discernible and consistent effect on indigenous Spanish scores.

While CONAFE was more effective for math instruction in primary school, CONAFE was more effective for Spanish instruction in secondary school. CONAFE appeared to increase *telesecundaria* Spanish scores by 0.8 to 3.4 points per year, while CONAFE increased math scores by 0.02 to 1.4 points per year. That effect represents an annual decrease in test score inequality between less-disadvantaged CONAFE and non-CONAFE students of 24 percent in math scores and 38 percent in Spanish scores.

Predictably, CONAFE students had significantly higher repetition and failure rates than did non-CONAFE students. CONAFE appeared to decrease repetition rates by an average of 0.13 percent per year, which represents an elimination of 6 percent of inequality in repetition rates between comparable CONAFE and non-CONAFE students. Overall, CONAFE decreased failure rates by 0.3 percent per year, but that effect was largest for the less-disadvantaged group.

These results imply that the Mexican government's provision of extra resources to particularly disadvantaged schools does result in measurable improvement in the learning of students in those schools. One should not draw from this result the conclusion that other

compensatory education will be equally effective. It may be that CONAFE's decentralized spending and management structure, its involvement of parents in local decision-making, its targeting scheme, or other unique aspects of its design cause CONAFE to be effective, and that those design aspects are absent from other compensatory programs.

Furthermore, at this stage we are unable to determine precisely which of CONAFE's interventions make it effective. Further research could help to draw lessons from CONAFE that would apply to similar programs in other countries. Therefore, future research using data on CONAFE support disaggregated by type of support should focus on determining what aspects of CONAFE's interventions make CONAFE effective. That research would allow broader generalization of our results to other countries.

Future research could also focus on the long-term effects of CONAFE intervention. Early research on the Head Start program in the United States found that Head Start increased test scores and decreased dropout and failure rates. Later research on Head Start, however, found that it had insignificant effects on college enrollment and adulthood earnings. We use test scores and short-term dropout and failure data as predictors of total schooling and lifetime earnings. But since the purpose of compensatory education is to help people escape poverty, the most telling evaluation should consider not merely whether CONAFE increases test scores, but whether CONAFE improves outcomes 10-15 years after a student attends a CONAFE school. Although CONAFE has existed for over 30 years, such evaluation may require panel data that is currently unavailable.

References

- Aughinbaugh, Alison. 2001. "Does Head Start Yield Long-Term Benefits?" *The Journal of Human Resources* 36(4): 641-65.
- Barnett, Steven. 1995. "Long-Term Effects of Early Childhood Programs on Cognitive and School Outcomes." *The Future of Children* 5(3): 25-50.
- Becker, Sascha O. and Andrea Ichino. 2002. "Estimation of average treatment effects based on propensity scores." *The Stata Journal* 2(4): 358-377.
- Castro, Claudio de Moura, Laurence Wolff and Norma García. 2001. "Mexico's Telesecundaria: Bringing Education by Television to Rural Areas." Working Paper, Inter-American Development Bank.
- Currie, Janet and Duncan Thomas. 1999. "Does Head Start Help Hispanic Children?" *Journal of Public Economics* 74(2): 235-62.
- _____. 2000. "School Quality and the Longer-term Effects of Head Start." *Journal of Human Resources* 35(4): 755-74.
- Currie, Janet and Matthew Neidell. 2003. "Getting inside the "Black Box" of Head Start Quality: What Matters and What Doesn't?" NBER Working Paper 10091. November.
- Dehejia, Rajeev H. and Sadek Wahba. 2002. "Propensity Score Matching Methods for Non-experimental Causal Studies." *The Review of Economics and Statistics* 84(1): 151-161.
- Garces, Eliana, Duncan Thomas and Janet Currie. 2002. "Longer-Term Effects of Head Start." *The American Economic Review* 92(4): 999-1012.
- Informes Estadísticos de CONAFE. n.d. "Primaria." Mimeo: Mexico City, México.
- Inter-American Development Bank. 1997. "Mexico: Distance Education Program." Report ME-0052.
- _____. 2003. "Mexico: Comprehensive Community Education Program." Report ME-0238.
- Jalan, Jyotsna and Martin Ravallion. 2003. "Estimating the Benefit Incidence of an Antipoverty Program by Propensity-Score Matching." *Journal of Business & Economic Statistics* 21(1): 19-30.
- Karly, Lynn A., Peter A. Greenwood, Susan S. Everingham, Jill Hoube, M. Rebecca Kilburn, Peter C. Rydell, Matthew Sanders and James Chiesa. 1998. "Investing in our children: What we

know and don't know about the costs and benefits of early childhood interventions." Santa Monica: RAND MR-898.

MINEDUC. 2000. "Evaluación del Programa de las 900 Escuelas." División de Educación General, Chilean Ministry of Education. Cited in Tokman.

Newman, John, Menno Pradhan, Laura B. Rawlings, Geert Ridder, Ramiro Coa and Jose Juis Evia. 2002. "An Impact Evaluation of Education, Health, and Water Supply Investments by the Bolivian Social Investment Fund." *World Bank Economic Review* 16(2): 241-274.

Paqueo, Vicente, and Gladys Lopez-Acevedo. 2003. "Supply-Side School Improvement and the Learning Achievement of the Poorest Children in Indigenous and Rural Schools: the Case of PARE." World Bank Policy Research Working Paper 3172, December 2003.

Park, Albert, Sangui Wang and Buobao Wu. 2002. "Regional poverty targeting in China." *Journal of Public Economics* 86: 123-153.

Pradhan, Menno and Laura B. Rawlings. "The impact and targeting of social infrastructure investments: lessons from the Nicaraguan social fund." *The World Bank Economic Review* 16(20): 275-295.

Rosenbaum, Paul R. and Donald B. Rubin. 1983. "The central role of the propensity score in observational studies for causal effects." *Biometrika*. 70(1): 41-55.

Secretaría de Educación Pública de México (SEP). n.d. Programa Nacional de Educación 2001-2006.

Slavin, Robert E. 1989. "Students at risk of school failure: the problem and its dimensions." In Robert E. Slavin, Nancy L. Karweit, and Nancy A. Madden, eds. *Effective Programs for Students at Risk*. Boston: Allyn and Bacon.

Tokman, Andrea. 2002. "Evaluation of the P-900 Program: A Targeted Education Program for Underperforming Schools." Working Paper 170, Central Bank of Chile.

World Bank. 1991. "Staff Appraisal Report: Mexico Primary Education Project." Report 9770-ME

_____. 1994. "Staff Appraisal Report: Mexico Secondary Education Project." Report No. 12529-ME.

_____. 1997. "Implementation Completion Report: Mexico Initial Education Project." Report No. 17192.

_____. 1998. "Project Appraisal Document for a Basic Education Development (PAREIB) Project." Report 17535-ME.

_____. 2002. "Project Appraisal Document on a Proposed Loan in the Amount of US\$300.0 Million to the United Mexican States for Basic Education Development Phase II." Report 23295-ME.