Central de Chile
Documentos de Trabajo

Central Bank of Chile
Working Papers

N° 170

Julio 2002

EVALUATION OF THE P900 PROGRAM: A TARGETED EDUCATION PROGRAM FOR UNDERPERFORMING SCHOOLS

Andrea Tokman

La serie de Documentos de Trabajo en versión PDF puede obtenerse gratis en la dirección electrónica: http://www.bcentral.cl/Estudios/DTBC/doctrab.htm. Existe la posibilidad de solicitar una copia impresa con un costo de $500 si es dentro de Chile y US$12 si es para fuera de Chile. Las solicitudes se pueden hacer por fax: (56-2) 6702231 o a través de correo electrónico: bcch@bcentral.cl.

Working Papers in PDF format can be downloaded free of charge from: http://www.bcentral.cl/Estudios/DTBC/doctrab.htm. Printed versions can be ordered individually for US$12 per copy (for orders inside Chile the charge is Ch$500.) Orders can be placed by fax: (56-2) 6702231 or e-mail: bcch@bcentral.cl.

BANCO CENTRAL DE CHILE

CENTRAL BANK OF CHILE

Documentos de Trabajo del Banco Central de Chile
Working Papers of the Central Bank of Chile
Huérfanos 1175, primer piso.
Teléfono: (56-2) 6702475   Fax: (56-2) 6702231

Documento de Trabajo
N° 170

Working Paper
N° 170

EVALUATION OF THE P900 PROGRAM: A TARGETED EDUCATION PROGRAM FOR UNDERPERFORMING SCHOOLS

Andrea Tokman
Economista Senior
Gerencia de Investigación Económica
Banco Central de Chile

Resumen
Políticas educacionales dirigidas a "malos" colegios son polémicas. Su concepto de discriminación positiva es discutible, y la falta de evaluación empírica seria, ha congelado el debate en un nivel excesivamente teórico. El mayor problema ha sido identificar efectos insesgados, dada la participación no aleatoria aun entre los colegios con mal rendimiento. Este trabajo contribuye al debate evaluando el Programa P900, estimando efectos libres del sesgo debido a características específicas (y no observadas) de cada escuela correlacionadas con su participación, considerando explícitamente la naturaleza cambiante del Programa y de su proceso de selección. Adicionalmente, prueba y rechaza supuestos simplificadores encontrados usualmente en la literatura, tales como efectos constantes del Programa, del sesgo y de las características fijas de los colegios, sembrando dudas con respecto a resultados anteriores. Se concluye que los colegios son seleccionados para participar en el P900 en forma compensatoria, sesgando las estimaciones hacia abajo. Más aún, el sesgo y el efecto del Programa aumentan en el tiempo. El efecto estimado de su versión de 1992 no es significativamente distinto de cero, mientras que es positivo para 1994 y 1996. Este último es significativamente mayor que el resto, por lo que efectivamente estaría ayudando a los colegios que participaron en él.

Abstract
Education policies targeted at the worst performing schools are controversial. Its positive discrimination nature has been debated, while the lack of serious empirical evaluation has frozen the discussion at a highly theoretical level. The biggest problem has been to identify unbiased effects, given non-random participation even within underperforming schools. This paper contributes to the debate by evaluating the P900 program - that provides material support to low-achieving Chilean schools- estimating effects that are free of bias due to unmeasured fixed school-specific effects, which are correlated with participation. It explicitly considers the changing nature of both the program and the selection process, computing yearly effects and biases. The simplifying assumptions usually found in the literature (e.g. constant program effects, bias, and school effects) are tested and rejected, thus shedding doubts on previous findings. The paper finds that schools were selected for the P900 in an unobserved compensatory manner, thus downward biasing uncontrolled estimates. Moreover, the bias and the estimated effects of the Program are increasing in time. The 1992 program is not significantly different from zero, while the 1994 and 1996 program´s are significantly positive. The latter is significantly higher than the rest, thus it would be effectively helping the schools that participate in it.

I am grateful to David Card and Kenneth Chay for their excellent advice throughout my doctoral dissertation at UC Berkeley, which led to this version of the paper. For their useful discussions, I thank Steven Raphael, Pilar Romaguera, Alejandra Mizala, Claudio Sapelli and Dante Contreras. Also the participants of seminars at UC Berkeley, University of Chile, Ilades/Georgetown, the Central Bank of Chile, and the North American Econometric Society Meeting. Finally, I would like to thank the authorities in Mideplan and Mineduc that gave me access to the data sets.
E-mail: atokman@bcentral.cl.

1. Introduction

Implementing an education program biased in favor of the worst performing schools is a controversial issue. It is difficult to justify directing resources to schools that appear to be providing education inefficiently, but several findings in the literature do. First, there is consensus in that the single most important determinant of student outcomes, and therefore school achievement, is socio-economic background. In essence, the low performance obtained by some schools might be reflecting low innate ability of their students and not inefficiency, since students coming from lower income/ education families will perform poorly, independently of the school they attend. Additionally, students pertaining to disadvantaged families are not free to move between schools, (not even in Chile were the voucher-type schools are majority). Thus, providing special support to the school where disadvantaged kids are locked in may be justified on equity grounds. Therefore, the combination of low ability and low mobility justifies the creation of education programs focused on the group of schools catering for the children of poor family background, over and above the support provided by the government to upgrade the quality and equity of education in general.

One of such programs is the Chilean program of the 900 schools (P900). It is a program targeted towards the 900[i] lowest-achieving subsidized schools measured by their 4th graders' score in the national standardized tests[ii]. It focuses on meeting the specific needs of the students, improve teacher quality, provide free textbooks and other educational material, and improve infrastructure. It does so by directly supplying the material and training required by the school. No cash is involved in the process.

The program was implemented for the first time in 1990 and is still active. It has remained within the Ministry's main education reform programs and has been amended continuously to better accommodate the specific needs of the targeted schools and the changing conditions in the Chilean society. All of this has been done based on the premise that the program has been effective in increasing the schools' test scores, but few evaluations exist, and more often than not, they are based on interviews and descriptive analyses of case studies (see for example, Carlson, 2000). Only the most recent evaluation (MINEDUC, 2000) -and probably the most complete- does some empirical analysis but is flawed in the sense that it compares schools that are different in many ways: some of them observable, for which the study does control (i.e. school size and private/public); and some unobservable, for which the study does not control. The results are probably biased if the unobserved and uncontrolled characteristics of the schools affect test scores regardless of whether they participated in the program or not, and at the same time have an effect on the probability of participating. For example, if participating schools are more entrepreneurial, the gain from the program is overstated when the resulting test scores are compared with "similar" nonparticipating schools, since those in the program would perform better even in the absence of the program. The objective of this paper is to re-evaluate the impact of the P900 program on test scores, taking into account such possibility.

The paper estimates a model that allows for the presence of unobserved fixed school characteristics that affect both participation and outcome. It explicitly compares the results and biases introduced by omitting relevant variables, and does so by imposing far less restrictions on the estimated parameters than do traditional fixed effects models. Relaxing the traditional restrictions on fixed coefficients seems desirable because the program possibly has a time-varying effect, as it has been continually under revision. In addition, the correlation between the unobserved relevant variables and participation may be time varying as selection guidelines are applied with different degrees of strictness in time, and therefore allowing them to vary (i.e. estimating time varying biases) also seems desirable.

The findings suggest that both restrictions imposed by the fixed-effect models are not supported by the data, and therefore there is a gain from estimating the less restricted model presented in this paper. In effect, the time varying coefficients estimated suggest that the Program has been increasingly effective in narrowing the achievement gap between participating and nonparticipating schools, and that selection based on unobserved school characteristics is compensatory and increasing, thus the uncorrected estimates are increasingly underestimated.

Although this paper contributes by advancing towards a less biased estimate of the program's effect on test scores, it does not claim to go all the way in estimating *THE* unbiased effect. In fact, several problems remain and should be dealt with to get closer to the real effect of the program. One of such problems–not considered in this paper–is that the unobserved heterogeneity of schools may be time varying. Another has to do with the fact that schools with lower test scores will tend to have bigger increases in test scores due to regression-to-the-mean type effects, and not due to the program itself.

The paper is organized as follows: first the program is described; then the possible evaluation strategies are discussed; the fourth section describes the empirical strategy followed; section five describes the data and the estimated results; finally, section six concludes.

2. The P900 Program

The nineties were marked with the desire to reform the Chilean education system, aiming at quality improvements with equity. This was a reaction to the observed schools' failure to achieve better student outcomes and a wide quality variance between schools, that left children from the least favored backgrounds (i.e. low parental education and income) with a very bad chance of ever overcoming their condition through education. Under this scenario, the *Programa de Mejoramiento de la Calidad de las Escuelas Básicas de Sectores Pobres* or *Programa de la 900 Escuelas* (Quality Enhancing Program for Elementary Schools at Lower Income Sectors or The 900 Schools Program) (P900) was born in the year 1990. Based on the principles of positive discrimination, it provides technical and material support (no cash) to the lowest decile in the fourth-graders' standardized tests distribution, which considers only subsidized schools[i].

The program focuses on meeting the specific needs of the learners in the underperforming schools required to increase their achievement, which is the final goal. It does so by providing technical and material support to teachers, directors and students in four main areas: i) improving teacher quality through periodical workshops where teachers can discuss both theoretical and practical matters associated with providing quality education that is coherent with the specific characteristics of the students and their environment. These workshops are tutored by trained supervisors that guide the discussion according to a plan designed by the central coordination division of the P900 program at the Ministry of Education (MINEDUC). It encourages teachers to adopt pedagogical practices that promote student participation, acquire responsibility for the students' achievements, and articulate the school-family link; ii) student workshops designed to help students at risk to raise their grades, and enhance their self esteem, social skills and creativity. In order to encourage participation, young monitors belonging to the same community supervise them. The students are chosen by the teachers and normally fall into one of two categories: those with a behavioral problem (e.g. the troublemaker, the absent-minded, the shy) and the underachievers (e.g. low grades and the like); iii) preparing and distributing textbooks, creating classroom libraries and educational material, and iv) improving school management through supervised work groups with representatives from the teachers body, directors and community authorities that design the institutional education program best fitted to the school´s objectives and the community's authenticity.

Schools may remain in the program for unlimited time. They "graduate" when they exceed the regional average test scores and/or when they win a *School Improvement Project (PME)*, which is another program of the MINEDUC that finances quality improving projects presented by the schools.

Even though some schools (most of them private) decided not to take part in the program in its earlier years (probably because of the stigma it carries), it is becoming more appealing with the years. Studies show that at the beginning some school officials were reluctant to join the program, worried about the responsibilities they would acquire with it, but as the program matured and became recognized, the initial lack of confidence faded out.

As for its organization, the P900 Coordination Unit depends on the Ministry of Education, and is responsible for the design and evaluation of the program's implementation by the regional offices. The regional supervisors, trained by the central coordination unit (directly during the first years and indirectly afterwards through the coordination director), train and coach school teachers, directors and monitors. They also participate in the teachers' workshops and management groups.

The regional officials are also responsible for selecting the participating schools, although their autonomy is restricted by the central unit's guidelines that establishes that schools in the lowest decile in the previous 4^{th} grade performance distribution must be picked. In essence, the regional authorities have had varying degrees of discretion in the selection process over the years and admit to having considered other elements in the decision, such as school indebtedness or personal evaluations.

Implementation is also done at the regional level and may vary between regions and across schools. For example, materials took a long time to reach certain locations in the early stages (some got to the schools with one semester delay), but efforts have been made to improve, apparently with success. (MINEDUC, 2000). Additionally, the quality and commitment of the supervisors is also important in the program's final outcome and this will vary across regions too.

The program has survived more than a decade mainly because of its dynamism and flexibility to adapt to the changes in the society and in the education system, thus learning by doing. Even though it has the same name and defined objectives today than in 1990, in practice, it is a different program every year.

In part its dynamism is due to its focalized nature, ideal for experimenting with new education policies that, when found effective, are included as part of the universal programs of the Ministry. Some examples are the supply of textbooks to all subsidized schools together with infrastructure improvements that resulted from the pilot implementation in the P900 schools.

Another modification refers to coverage. Until 1997, only primary schools could enter the program based on their fourth grade standardized test scores and the vulnerability of its students. Starting in 1998, the program was extended to include pre-school and high school. Eighth graders' standardized test scores were also considered in the assignment process.

Until 1998, participating schools had no explicit responsibility, even though they had to *work* (i.e. provide time and teachers) for the program. Starting in 1998, schools were asked to assume a written commitment on what the program would require from them. Additionally, schools were mandated to remain in the program for a minimum of three years. That same year, a new line of action was added to the program that involved working jointly with the families.

The program has an approximated annual cost of US$2.6 million. It was financed through international cooperation from the Swedish and Danish governments in 1990-1991 and forms part of the national budget since 1992. The cost of the program accounts for less than 1% of the total education budget and around 9% of the budget in primary education, and did not vary much in time until its coverage was extended in 1998. The program increases government spending in elementary education by less than 2% per student.

3. Evaluation of the Program's Effect

In theory, given the design of the selection guidelines, the evaluation of the P900 should be straightforward. The way the program is presented to the public and the researcher makes it clear that schools are selected on the basis of their previous fourth grade average test score, and since such information is public, a simple comparison of the schools in and out of the program, in the margin, should be enough to identify the program's effect on test scores, at least for the schools that have similar characteristics to those in the cutoff area. This appreciation is correct only if selection is based exclusively on test scores and no other school feature that might influence achievement is considered. Unfortunately, as explained earlier and later corroborated in the data section, this is not the case. The guidelines are followed loosely and selection is based, in part, on characteristics unobserved by the researcher. In fact, regional program administrators recognize that other elements were considered in the decision, such as school indebtedness and teacher stability. Therefore, because these characteristics are believed to affect the ability of the school to obtain a high-test score, leaving them out will lead to an omitted variable bias in the estimated program effects. In particular, if schools were selected in a (unobserved) compensatory manner, the uncontrolled comparison would lead to a downward biased estimate of the real impact of the program since those schools would achieve lower scores in the exams even after controlling for observed characteristics. This is what happens, for example, when a larger debt limits the school´s ability to perform well in the standardized tests and the selection process favors the more indebted schools. On the other hand, positive unobserved selection would lead to an upward biased estimated effect. This would be the case if authorities selected schools based on their evaluation of which schools would benefit the most from the program. If that happens, then it is not surprising that the program has a bigger impact on test scores than if a random school was chosen. Thus, the estimated effect does not represent the expected effect for every school.

The endogeneity problem mentioned above is a common occurrence in program evaluation, and has been dealt with in previous literature in several ways. Probably the most popular one consists in the use of instrumental variables. In our case it would require us to find some school feature that is associated with participating in the program but not with the school's ability to perform well in the tests. Finding such a variable is quite difficult in this scenario, and using one that does not completely satisfy the restrictions leads to even bigger biases (Bound and Jagger, 1995). For example, this would happen if using proximity to regional centers as an instrumental variable for participation, believing that the schools that are located near the selecting authorities can lobby and improve their chances of being picked. But at the same time, schools situated in the main regional cities (and therefore near the authorities) tend to be endowed with better teachers (it is easier to attract good teachers to bigger towns) and students (bigger towns concentrate wealthier families), and therefore will perform better, *ceteris paribus*. Therefore, since there are no obvious instrumental variables for our specific problem, this methodology will not be applied herein.

A second strategy commonly used in the literature to correct the omitted variable bias that rises because of non random selection, consists in adjusting the estimated effects by using Heckman type selection correction models. This methodology estimates a first stage probit equation with maximum likelihood for the probability of selection/ participation and uses the estimated parameters (using the properties of a normal probability distribution function) to construct a selection correction term (normally known as the Mills ratio) for each observation. The second stage regression is estimated using the correction term as an additional regressor to obtain unbiased estimated effects. Even though the procedure is easily computable and has been widely applied to the analysis of education and a host of other settings, the high sensitivity of the results to the distributional assumption - its reliance on the normality assumption used to compute the selectivity term- has drawn the attention of empirical economists, and the generality and acceptability of the method has been questioned. In the case of our specific problem in mind (i.e. the evaluation of the P900), we cannot guarantee that the joint normal distribution of error terms in the selection and outcome regressions is satisfied and therefore will not use this method either.

This paper deals with the selection problem by estimating fixed-effects -type models that assume that school heterogeneity can be captured by a fixed school-specific term. In the case at hand, this would mean that the unobserved features of the schools that are considered in the selection process are time invariant. This is true, for example, if the schools´ proximity to authorities is determinant in their participation, because unless schools or authorities move away, this characteristic does not change in time. Other school characteristics that may influence selection, such as stability of teachers at the school, vulnerability of enrolled students, or propensity to run deficits, are probably time invariant too.

Obviously, one can think of time varying characteristics that may be considered in the selection process and affect school outcomes as well. This paper does not claim that it is not possible, but chooses to ignore them to simplify the analysis. Thus, the effects estimated herein are free of omitted variable bias due to the presence of unmeasured *fixed* school specific effects that are correlated with participation and outcome variables. Biases due to variable school effects are not considered.

Also, several restrictions imposed by the fixed effect model seem to be unacceptable for the P900. In particular, fixed effect models estimate constant coefficients in time, which means that the estimated treatment effect is assumed to be the same every year. This is probably not desirable in the P900 case, given the continuous changes the program has undergone in terms of resources devoted to it, its application, coverage, and the like. In essence, it is a different program every year, and one would expect different effects over time. Additionally, the fixed effect model assumes a constant bias due to nonrandom selection of schools based on fixed characteristics. In our case, it is probably true that as the program evolved and the authorities responsible for selection changed in time, the arbitrariness in the selection process did so too. For example, it can be expected that different regional authorities consider different variables in the decision, or that new central coordinators value discretion differently. In fact, interviews show that regional authorities are increasingly feeling left aside, since in the earlier years they had some influence in the selection decision, whereas today such impact is null (MINEDUC, 2000).

Thus, the traditionally accepted restrictions imposed by the fixed effect models are not adequate for this particular program and a less restrictive model seems necessary. The paper estimates one of such models that allow biases and effects to vary in time. The correlation of the fixed effects and the participation variable are explicitly considered, in order to be able to estimate if and how the assignment discretion has changed in time, and what its impact on the uncorrected estimated time varying effect is. Moreover, the traditional fixed effect model can be nested in the general model and its restrictions tested in order to evaluate its adequacy for the problem at hand.

4. The Model[iii]

As with education production functions, we assume that the outcome measure (i.e. test score) of school s in time t (Y_{st}) is correlated with a set of fixed observed (F_s) and unobserved characteristics (C_s), variable school characteristics (X_{st}), and a yearly program participation dummy ($P900_t$), as described in equation 1.

$$(1)\, Y_{st} = j'_t F_s + a'_t X_{st} + b'_t P900_{st} + g'_t C_s + e_{st}$$

where the fixed unobserved school effect (C_s) is uncorrelated with the error term (ε_{st}) but possibly correlated with the other fixed and varying characteristics. In particular, if schools are selected in a compensatory manner (though not observed by the researcher), then those schools will expectedly perform worse than average even in the absence of the program and, therefore, a negative covariance between the unobserved school effect and the participation dummy will arise, biasing the estimate treatment effect down.

In general, since the school characteristics (C_s) considered by the selecting authorities are fixed, they probably affect selection each year. Thus, if the unobserved fixed school characteristics are correlated with the P900 participation dummy, it may also be correlated to its leads and lags, as expressed in (2)[iv]. (Only three time-periods are considered because of the data available).

$$(2)\, C_s = l_1 P900_{s1} + l_2 P900_{s2} + l_3 P900_{s3} + x_s = l' P900_s + x_s$$

Note that the *lambdas* are allowed to vary by year, thus allowing a varying influence of the unobserved characteristics in the selection process every year, to keep in line with the possibility that the application of the guidelines, the selection authorities, or the elements considered by them change in time, thus allowing for a bias in the estimate effects that may vary in time.

Substituting C_s into the school production function (1) we obtain a model based on observable fixed and varying school characteristics and present, past and future program participation dummies:

$$(3)\ Y_{st} = j'_t F_s + a'_t X_{st} + b'_t P900_{st} + g_t l' P900_s + (g_t x_s + e_{st})$$

which is a restricted specification of the following reduced form unrestricted model:

$$(4)\ Y_s = \Phi X_s + \Pi P900_s + e_s$$

where $Y_s=(Y_{s1}, Y_{s2}, Y_{s3})'$, $P900_s=(P900_{s1}, P900_{s2}, P900_{s3})$ and $e_s=(e_1, e_2, e_3)$.

The restrictions imposed to the unrestricted model in (4) by our least restricted model (3) imply the following nonlinear restrictions on its parameters:

$$(5)\ \tilde{\Pi} = \begin{bmatrix} b_1 + g_1 l_1 & g_1 l_2 & g_1 l_3 \\ g_2 l_1 & b_2 + g_2 l_2 & g_2 l_3 \\ g_3 l_1 & g_3 l_2 & b_3 + g_3 l_3 \end{bmatrix}$$

that can be tested using minimum distance estimators (Chamberlain 1982) to evaluate the adequacy of such restriction on the existing data.

Moreover, using restricted GLS, estimates of the effect of the program in test scores (β's), the effects of the fixed unobserved school characteristics in test scores (γ's) and the correlation between participation in the P900 and the unobserved fixed characteristics (λ's) can be obtained for different points in time. The increased flexibility of the model as compared to a model with fixed coefficients (i.e. traditional fixed effect models) seams desirable –as argued earlier- since the program has evolved and consequently its effect and biases probably have changed over the years.

Additionally, in order to advance towards a simpler yet acceptable model for the data, restrictions on the estimated parameters can be tested against the unrestricted or least restricted models to check whether they are satisfied. For example, if one would like to check the adequacy of the fixed effects model's restrictions in the P900 scenario, one would have to jointly test the assumptions that the treatment effect (β) and the effect of the unobserved school characteristics on test scores (γ) are the same every year. This can be done by testing the following restrictions:

$$(6)\quad \tilde{\Pi} = \begin{bmatrix} b + l_1 & l_2 & l_3 \\ l_1 & b + l_2 & l_2 \\ l_1 & l_2 & b + l_3 \end{bmatrix}$$

The rest of the paper will do exactly that: It will estimate the most unrestricted model and test the restrictions imposed by alternative models using traditional specification tests. The traditional fixed effects results will be compared with less restrictive models and the acceptability of its assumptions will be evaluated. Results will also be compared to the cross section estimates, and explicit estimations of the bias will be calculated.

5. Empirical Analysis

5.1 The Data

The data set used was obtained by merging several school level yearly data sets provided by the Chilean Education Ministry, that contain average school (math and Spanish) 4th grade standardized test scores for 1988, 1992, 1994, 1996; P900 participation dummies for 1990-1996; economic vulnerability index[v] for 1990, 1992, 1993 and 1996; and 12 regional dummies. Unfortunately, at the time the research was done, the test scores for 1990 were not available. Additionally, because big changes were introduced in the program in 1998 that delayed the 4th grade test to 1999 due to the inclusion of a 10th grade test, the program's effects will be calculated for the schools that participated in 1992, 1994 and 1996.

The sample was restricted to public schools to avoid dealing with private schools that declined to participate in the program. Such restriction to the sample should not matter much because over the past 10 years, more than 80% of the participating schools have been public. Besides, their probability of not accepting the program if selected is almost zero, since they depend directly on the selecting authorities. Thus, the findings in this paper will be interpreted as the effect of the P900 in public schools, but will say nothing about the effects on private subsidized schools.

The characteristics of schools in and out of the program can be seen in Table 1 and Graph 1. On average, there are 3,600 non-participating schools and 530 participating schools. The data show that, on average, participants have lower previous test scores than non-participants, which is consistent with the positive discrimination objective based on previous test scores. Additionally, the gap between test scores of participating and non-participating schools tends to decrease and even disappear in time, and could be interpreted as showing a positive impact of the P900 on test scores. Unfortunately, such an observation may not coincide with reality because the comparison group is not a good counterfactual for what the test scores would have been had they not participated. This is so because participants are different from non-participants in both observed and unobserved ways.

If the selection process was conducted as explained in the guidelines, considering only previous test scores, then only the lowest achieving schools would take part in the program. But this is not the case, schools with very low-test scores not participating as well as high achieving schools participating can be observed.

Furthermore, given that the intention of the program, as expressed by the authorities, is to correct the achievement deficit of schools that have the most vulnerable students, one would expect that the selecting authorities would consider student vulnerability in the process too. Even so, participation, controlling for economic vulnerability and test scores, is not totally predictable either. In particular, if eligibility to the program were determined only by pre-test scores and economic vulnerability, the plots presented in graph 2 -which marks the participating schools with a plus sign in a test by vulnerability space- would show all participating schools in the lower right area and non participants further to the left and up. This does not appear to be the case. Nonparticipating schools with lower test scores and higher vulnerability than participating schools are observed. Some schools that participate have high-test scores and low vulnerability, and therefore should not be in the program. This is also true when the analysis is done at the regional level.

As described earlier, if the factors that explain such divergence between predicted and actual participation are correlated with the schools' ability to obtain higher test scores, and the researcher does not observe them, estimating the effect of the P900 in test scores by comparing participating and non-participating schools will be biased.

Another look at the data through probit regressions confirms the existence of selection not based strictly on the variables mentioned in the guidelines. Again, vulnerability is included in accordance with the program's objective even though the guidelines don't mention it. Table 2 presents the results of yearly regressions for the probability of participating as a function of previous test scores, previous economic vulnerability index and regional dummies. As expected, every year schools with lower previous test scores have a higher probability of participating in the program (e.g. a one point increase in test scores reduces the probability of participating in the program by 0.09 times the standard deviation in 1992). Also, for 1994 and 1996, we find that the schools with the most vulnerable kids are more likely to participate.

The probability of correctly predicting participation is between 62 and 69%, thus indicating the presence of unexplained participation in 31 to 38% of the times. If the fraction of participation that cannot be explained by the observed variables is randomly assigned between schools, it does not present a problem. But if it depends on that have an effect on test outcomes, then uncontrolled comparisons between participants and non-participants will provide biased estimates of the program's effects. This paper deals with such possibility by explicitly estimating the correlation of the participation indicator with fixed unobserved school characteristics. Additionally, since it appears that not all years are equally predictable nor is the selection process equally strict in observing the selection guidelines, we will allow for a varying correlation/bias in time. This suggests that selection based on other school characteristics not observed by the researcher, but fixed in time, may be different from one year to the next, probably because of more or less discretion allowed or more or less weight allocated to each characteristic in the selection process. Also, selection may vary because the program started alone but in time several other ministry programs became popular and probably changed the desire/possibility of some schools to participate in the P900. This would be the case, for example, of the education enhancing projects *Proyectos de Mejoramiento Educacional* (PME) that became well known during the second half of the decade (e.g. in 1997, more than 80% of the schools had a PME) and restricted the sample of schools available to participate in P900, because participation in both programs is not allowed.

5.2 Cross-Section Estimates

The first set of estimates shown in Table 3 consists of the traditional cross section OLS estimates that ignore the time pattern of the data and unobserved school effects in the analysis. This is what the most naïve researcher would do and therefore will be used as a benchmark. All the regressions control for current and lagged vulnerability, lagged test scores, regional dummies and total previous years in the P900. The 1992 equation differs from the other two by controlling for the 1988 test instead of the 1990 test, due to the lack of data for that year.

The estimated effects of the program appear to be substantially different every year, ranging from significantly negative in 1992 to positive in 1996. Taking these estimates at face value, we would conclude that the program initially had a detrimental impact on participating schools, but as it matured it became effective in improving the schools' relative test score. Such conclusions would be wrong if schools were selected on the basis of their unobserved ability to produce higher/lower test scores, since the estimates are biased. In addition, if the program executors, eligibility guidelines and the strictness of its application change in time, the bias may also change in size and direction, impeding us to infer the real yearly effect or even whether the estimates are lower or upper bounds without further analyzing selection every year.

Assuming no omitted variables bias, the estimated effect of the P900 in 1992, 1994 and 1996 (i.e. initially negative and then positive), may be implying that program participation takes time to impact test scores. Probably, as the program matured and was included as a regular ministry program, it became more effective in generating a faster improvement in test scores[vi]. The second part of Table 4 includes lagged participation dummies to capture the possibility of such delay in the program's impact. The coefficients reported are for the current and lagged participation dummies. The immediate or current effects are presented in the first row and are conceptually equivalent to those presented earlier (i.e. the effect on test scores in 1992 of having participated in the P900 that year, and controlling for vulnerability and length of participation is –2.09). The signs and significance remain.

The diagonal elements show the impact of the 1992 P900 in 1992, 1994 and 1996 test scores. It does not appear as if the program's positive effect just took time to show up, since the coefficients are negative in 1992 and 1994 and null in 1996. It could be interpreted as a negative effect of having participated in the 1992 program that gradually vanished four years later. In contrast, the negative estimated effect could be interpreted as driven by the underlying characteristics of the schools selected that would have performed even worse without the program. The diagonal for the 1994 program presents a positive current effect and no effect in 1996, possibly implying that the effects hold only in the short run.

In sum, the cross-section regressions suggest that the program has a different impact on test scores, depending on the year analyzed. The 1992 program appears to be detrimental to the schools' test scores, whereas the 1994 and 1996 programs appear to succeed in increasing the relative test scores of the affected schools. These results could be the outcome of the evolution of the program that has been modified to better serve the participating schools, but it could also be masking the real outcomes by not including the biases that arise from the non-random selection process. Such biases may be varying in time and therefore they may not even reflect an upper or lower bound for each year, misleading the reader to the wrong conclusions. The following section uses panel data and explicit assumptions on the correlation between the participation dummy and the fixed unobserved school variables to explore these possibilities.

5.3 Panel Data Estimates

This section uses panel data to estimate the least restrictive model described in section 3. The model includes all leads and lags of the P900 dummy variable (i.e. equation 4) that come from assuming that if selection into the program in a specific year considers fixed unobserved characteristics of the schools, then it will probably consider them in other years too (but not necessarily with the same weights). The unrestricted model serves as a benchmark, although it does not allow the independent identification of the effect of the program every year, or the effect of the unobserved school characteristics in test scores and participation. Therefore, a second estimation obtained using restricted GLS includes some restrictions that permit the identification the individual effects. These restrictions can be tested against the unrestricted model, and its adequacy, given the data, can be evaluated. Furthermore, other models that include additional restrictions are estimated and its restrictions tested in an attempt to find a simple model that describes the data acceptably.

Table 4 reports the unrestricted π matrix (i.e. the coefficients on the leads and lags of the P900 dummy for each test year). The model also includes current and previous vulnerability, previous test score, total previous participation years, and regional dummies[vii]. The reported standard errors are heteroskedastically consistent. The unrestricted impact effects (i.e. same year effects) follow the same pattern as those in the cross-section regressions (see the diagonal elements of Table 4): negative for 1992; positive and increasing for 1994 and 1996 (all significant at 99%). Once again the findings suggest that initially the implementation of the program was deficient, but as the program matured it was modified and improved, thus reporting an increased relative test score for the participating schools. But, as stated earlier, this interpretation is not necessarily true because we may be comparing schools that are inherently different, whose achievement differs from the average expected achievement (controlling for observable characteristics), even in the absence of the program. If such inherent differences between schools that cause the bias in the estimated coefficients are constant in time (i.e. proximity, entrepreneurialship), we can use the model described in section 3 to explicitly estimate it. Additionally, individual identification of the two elements that determine the size of the bias -the effect of the unobserved characteristics in test scores and participation probability- can be estimated.

Furthermore, from the analysis of the estimated unrestricted coefficients presented in Table 4, it appears that the previous explicit assumption on the correlation of the unobserved school characteristics with all P900 participation dummies is not far from reality. This is evident from the non-zero off diagonal elements, which may be zero if previous participation does not affect current test scores.

Table 5 presents the estimated program effects of the successively restricted models. The first model is the least restrictive one. It assumes the existence of an explicit form for the correlation of the school effect and the participation dummies, but allows complete flexibility in the coefficient estimates in time for all variables including the effects of the program, the school fixed effects and the correlation terms.[viii] Different effects in time are expected, given the dynamic nature of the program design, and therefore allowing for different β's is desirable. As explained earlier, the program has been active for more than a decade and has undergone several revisions and amendments that suggest that the actual effect of the program in relative test scores may not be the same depending on the year analyzed, since in essence the program was different every time.

Also, the size and sign of the omitted variable bias is jointly determined by the effect of fixed school characteristics on test scores and participation probabilities (γ and λ). Once more, it is possible that both of these elements vary in time. For example, tests may emphasize different subjects in different years, and therefore could be affected by the unobserved school characteristics in different ways. And, given the varying degree of influence that the selectors have in time, the correlation of the unobserved characteristics with the probability of participating may vary too. For example, some selectors may find it desirable to provide the program to those in the most isolated areas that will not be reached by any other program, whereas others may value proximity to distribution centers to make sure that the help reaches them promptly. In any event, not allowing the γ´s and ë´s to vary may be unacceptable given the nature of the program and may be seriously affecting the results. Fortunately, the framework allows us to analyze such restrictions critically and test their validity given the data.

No evidence exists against the restrictions imposed by our most general model described in equation (3) or restriction (5) (i.e. $\chi^2(1)=0.06$ is not rejected)[ix]. The estimated effect of the 1992 P900 (β_1) is not statistically different from zero. The effects for both the 1994 and 1996 programs are positive and significant ($\beta_2=3.90$ and $\beta_3=9.34$, respectively). The interpretation of the results would imply that the program in 1992 had no effect on the achievement of its participants –contrary to what the cross-section results indicate- but positive and increasing effects for the program in later years, implying that participation in 1992 produces a test improvement of 3.9 points, which is tripled by schools that participate in 1996. Given that the scores grow on average for all schools (hopefully because of the success of the education reform), this impressively higher effect of the latter programs may not be explained by the mean reversion of the scores as some might think.

In terms of the estimated bias due to the presence of unobserved fixed school characteristics, its component are as follows: the estimated effects on achievement are positive and significant, while those on participation are negative and significant. This means that the unobserved characteristics tend to be achievement increasing, and that selection is done choosing schools that would perform worse than the rest after controlling for all observable characteristics.

The above findings suggests that every year sorting/selection into the program is compensatory (i.e. negative λ's). That is, schools that participate in the program have lower unobserved ability, conditional on student vulnerability, previous test scores, and regional distribution. Such negative selection implies that comparisons of schools in and out of the program are misleading because the participating schools would have underperformed relative to the other schools had they not participated in the program. In other words, the uncontrolled effects are underestimated (i.e. biased downward). Moreover, the bias increases every year, suggesting that compensatory selection has grown with the program's successive implementations. One explanation for increasingly selecting (unobserved) disadvantaged schools is that, at the beginning, schools did not have much choice on how to receive help to improve their quality, but as the other programs of the ministry became popular, the most entrepreneur underperforming schools participate in the alternative programs and therefore the sample of schools available for P900 participation is reduced to the ones that did not win a PME.

A common assumption made when evaluating a program like the P900 is to estimate *THE* treatment effect assuming that the impact on test scores is independent of the year it was implemented. Such restriction may seem reasonable in some cases, but apparently not in the case at hand, since the program has evolved continuously and is quite different today from its first version, thus assuming that its effects are the same is unrealistic. If the data reject this assumption, then the models that estimate one effect would be actually presenting an average yearly effect, which states nothing about each year's effect or whether it has improved or worsened in time. This kind of information is crucial if we want to use it to modify the design of the program.

Part II of Table 5 estimates a model with constant effect for the P900. That is, it assumes that the P900 in 1992, 1994 and 1996 are equivalent in terms of how much they help the schools in their relative performance. The estimated program effect is positive and significant (β_2=3.21) but, as expected, the model is rejected when compared to the previous less restrictive model and to the fully unrestricted one. Thus it confirms the intuition that the program is different every year, in terms of the resources they provide, as well as the way they provide them, and therefore each yearly program must be considered as an independent unit.

Another typical assumption in this kind of papers is that the unobserved school characteristics have the same effect on test scores, regardless of the year. Again, most of the times such restriction is not tested, and the implications of imposing it into data that do not accept it are not analyzed. Part III estimates a model in which the school fixed effect is constant (and equal to one). Such model is not rejected by the data when compared to the fully unrestricted model, but is rejected when compared to the least restrictive model in Part I of Table 5. The estimated program effects are similar to those estimated in our less restrictive model, but smaller in magnitude. The correlation coefficient is now larger and significant at 99% confidence every year.

Additionally, it is not uncommon to assume a fixed correlation between the unobserved fixed variables and the P900 participation dummies, which implies that selection based on unobserved fixed characteristics remains unchanged in time. The data reject such restrictions when compared to the least restrictive model (i.e. additional $\chi(2)$ is 7.05), implying that even if selection based on unobserved fixed school characteristics is compensatory every year, it is not the same for all years and thus cannot be estimated as such. The estimated program effects have the same signs as those obtained above, but their dispersion is less because they are now corrected by the average bias, which is low in 1996 and high in 1992 when compared to the yearly bias.

Finally, researchers traditionally estimate a fixed effects model that assumes constant program effects over time (which is rejected by our data) and constant school effects (also rejected by the data). The results of imposing both restrictions in the data are presented in Part V. The estimated constant program effect is positive and significant (3.79); the correlation coefficients are negative and significant. The underlying conclusion is that the program is efficient in improving test scores. Unfortunately, the studies do not test the validity of such results in terms of the underlying restrictions. If they did, they would find evidence (as we did) against their restrictive model that would shed doubts on the simplified conclusion, especially because it would look as if the program in 1992 had a positive impact in test scores while there was actually no effect, and that the effect for the 1996 program is lower than what the unrestricted model suggests and the data accept.

The last set of results in Table 5 presents the traditional complete fixed effects model that would be obtained from any panel data estimation software. The additional restrictions are constant coefficient for all other variables in the model. Obviously, the model is rejected. Still, the estimated program effect is positive and significant.

In sum, we find that most of the simplifying assumptions made in traditional program evaluations and included as restrictions to our unrestricted model are rejected by the data. Imposing them in the estimation leads to a mistaken estimated effect of the program analyzed. The restrictions imposed by the first model are not rejected and, consequently, it appears to be the most adequate model to describe the data. Its results suggests that schools selected to participate in the program have a lower unobserved ability to obtain higher test scores than similar schools (i.e. schools with similar vulnerability, region, previous test scores, etc). In other words, schools are selected in a compensatory manner, and thus the estimated effects that do not control for such selection underestimate the real effects. Moreover, we find that selection is significantly different and increasing over time, which could conform to a change in the selectors' objective function, the guidelines or the sample of schools available for the program every year considering the introduction of alternative programs throughout the decade.

The results also suggest that the impact of the program is different every year. When controlling for the bias that arises from selection based on fixed unobserved characteristics, the 1992 P900 appears to have had no effect on test scores, whereas the 1994 and 1996 programs appear to have been effective in increasing the test scores of participating schools. Moreover, the effect is three times as big in 1996, thus suggesting that the program has been revised properly and its operation has improved to become a successful program today.

6. Conclusions

The paper evaluates an education program directed to underperforming schools in operation in Chile since 1990. Its positive discrimination nature has lead to both adherents and adversaries and the lack of serious empirical evaluations has left the debate at a highly theoretical level. The debate has centered on totally uncontrolled comparisons across schools in and out of the program, cross-section regression type analysis that control for very few observed school characteristics, and restricted panel data analysis, all of which report erroneous program effects if the selection of participating schools is non-randomly assigned (and not controlled for explicitly).

The Ministry's website and media communications argue that the program is highly effective and show a table of the changes in the average test scores of participating schools as compared to the rest of the subsidized schools during the 1990´s. They claim that the bigger percentage increase in test scores of participants is a clear sign of a successful program but the comparison is misleading, because such behavior is naturally expected to happen as scores revert to the mean. Moreover, it says nothing about how the selected schools would have performed without the program, since they differ systematically from the rest of the schools both in observed and unobserved ways.

Cross-section regression analysis is better, because at least it allows for comparisons between schools that are similar in observed characteristics. This would suffice if the schools in the program were selected based only on the features observed by the researcher. Unfortunately, contradicting the program's design, unobserved characteristics are considered in the selection process and are also important determinant of school outcomes, thus rendering biased cross-section estimates.

This paper contributes to the discussion by estimating the effect of the P900 free of omitted variable bias arising from the presence of unmeasured fixed school effects, which are correlated with participation. It estimates the least restrictive model that explicitly considers the ever-changing nature of both the program and the selection process. Such model allows program effects and biases to vary in time. It also tests and rejects the simplifying assumptions usually made in the literature (e.g. equal program effect all years, equal bias, equal school fixed effect).

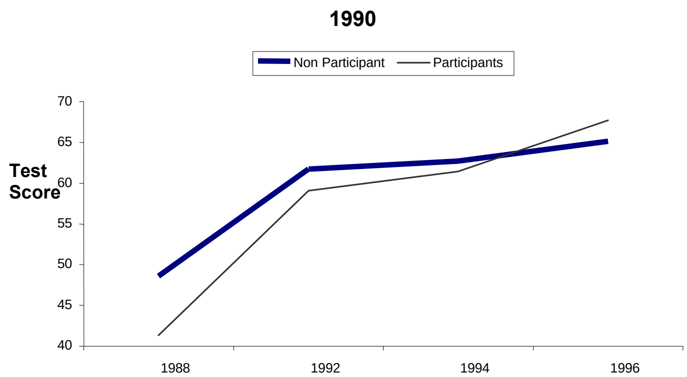
The main results suggest that schools that have participated in the P900 were selected in a compensatory manner, even after controlling for vulnerability and previous test scores. In other words, the selected schools are less likely to achieve a high-test score than schools with similar observed characteristics, thus leading to downward biased uncontrolled estimates. Moreover, the bias is increasing in time. Some possible explanations are included in the paper.

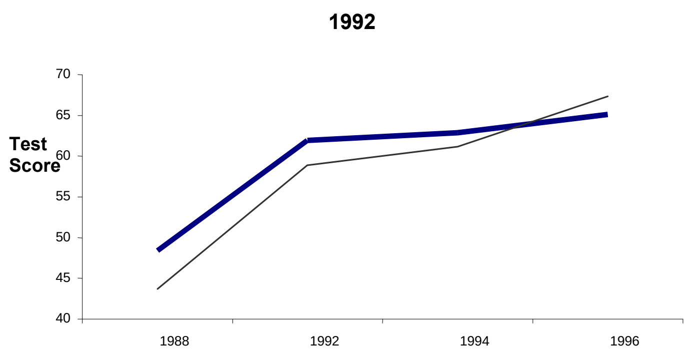
Additionally, the estimated effect of the P900 in relative test scores is different every year and has increased in efficiency as the program has become consolidated in the regular MINEDUC programs. The estimated effect for the 1992 program is not significantly different from zero, as opposed to negative uncontrolled effects. The effects for 1994 and 1996 are significantly positive and higher than the traditional cross-section estimates. The effect for the 1996 program is significantly higher than the previous ones, thus effectively helping the schools that participate in it.

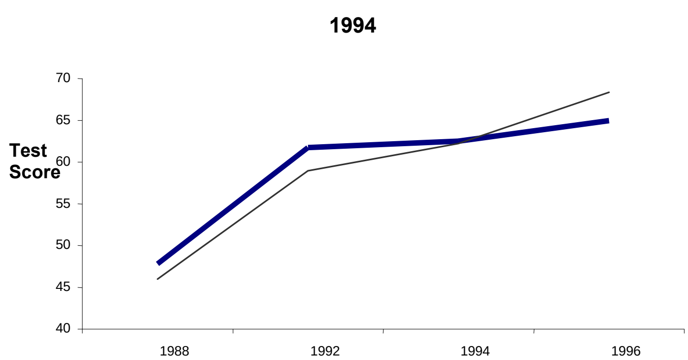
It is important to keep in mind that the estimated effects are not totally unbiased. In fact, the procedure used only corrects for biases that arise due to the presence of *fixed* unobserved school characteristics that have an effect on both school outcomes and participation. It does not claim to correct for biases that arise from time varying characteristics considered in the selection process. The latter is left open for future research.

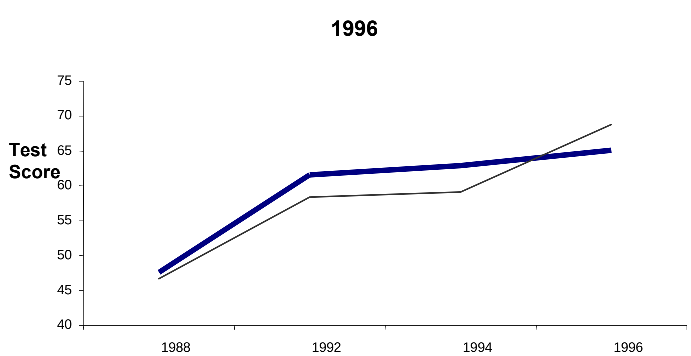
To conclude, the program has proven effective in narrowing achievement gaps. A learning process in the implementation allowed for an increased efficiency. The researcher is tempted to conclude with one final suggestion. While the flexibility given to the regional secretaries to select the schools to be benefitted by the program permits them to adapt to local and school specificities not incorporated in the general selection criteria, the Ministry should request and tabulate the main arguments to include or exclude schools that, meeting the general selection criteria, were not program beneficiaries. This would allow for better analyses of performance in future studies and guide policy design correctly.

Graph #1
Test Scores by Yearly Participation



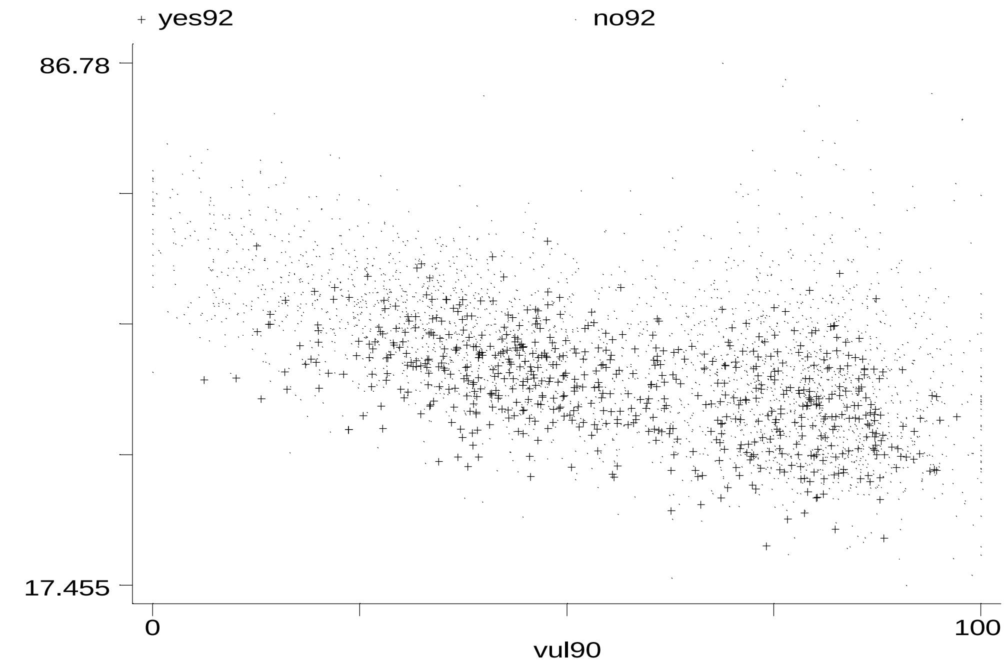




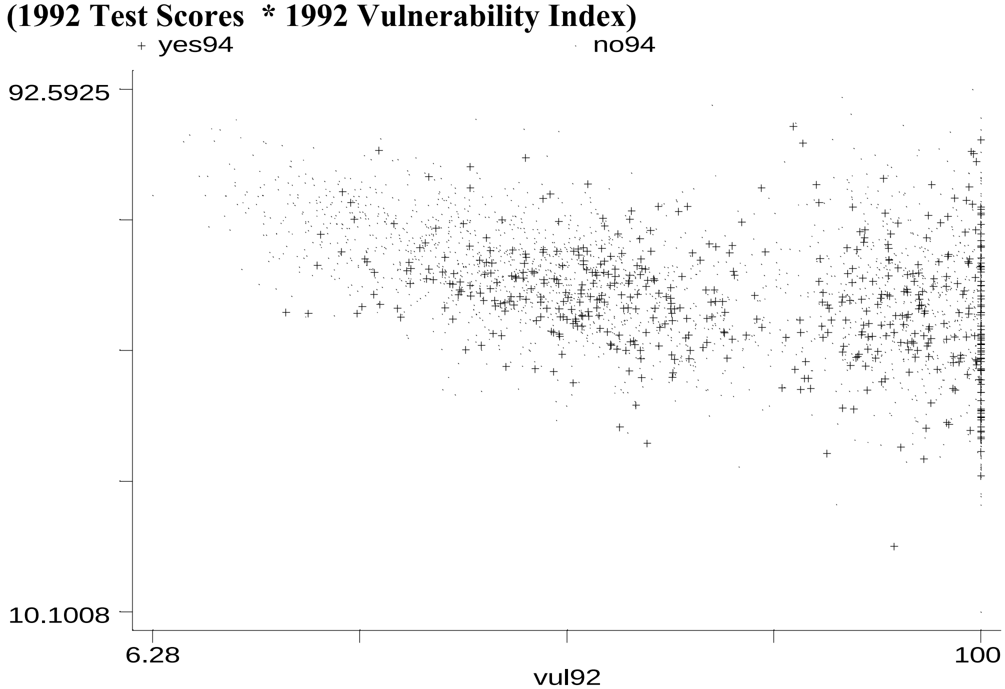


Graph #2 P900 Selection by Year (Test $_{t-1}$ * Vulnerability $_{t-1}$)

P900 Participation in 1992
(1990 Test Scores * 1990 Vulnerability Index)



P900 Participation in 1994
(1992 Test Scores * 1992 Vulnerability Index)



P900 Participation in 1996
(1994 Test Scores * 1993 Vulnerability Index)

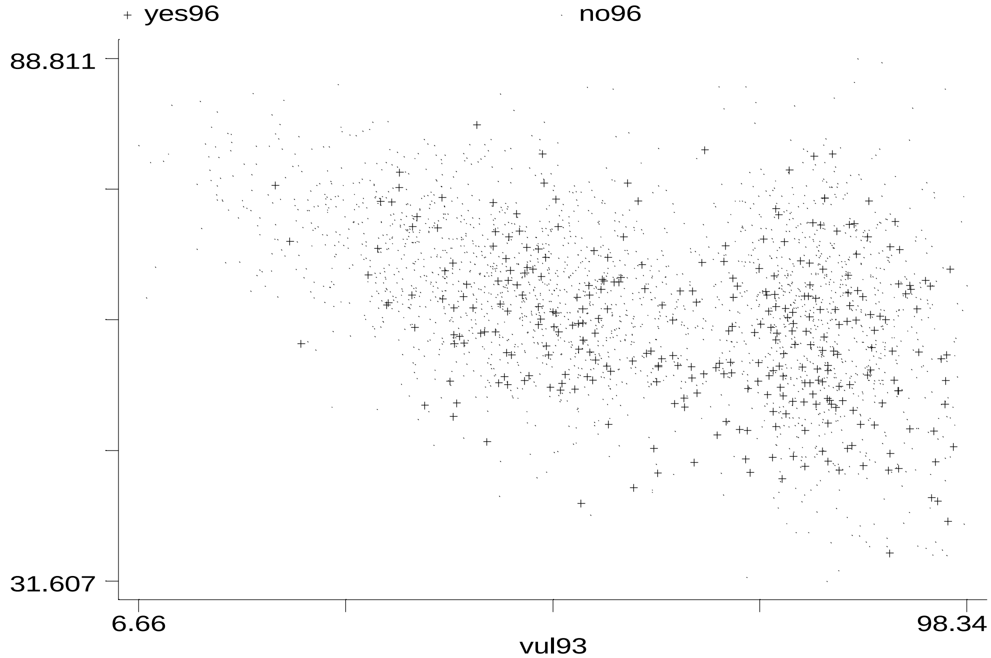


Table 1
Means by Yearly Participation in P900 Program

	1990				1992				1994				1996			
	NO		YES		NO		YES		NO		YES		NO		YES	
Variable	**Mean**	**Std Dev**	**Mean**	**Std Dev**	**Mean**	**Std Dev**	**Mean**	**Std Dev**	**Mean**	**Std Dev**	**Mean**	**Std Dev**	**Mean**	**Std Dev**	**Mean**	**Std Dev**
Number of Schools	3625		524		3485		664		3563		584		3785		364	
Test 1988	48.58	9.96	41.33	5.61	48.38	10.15	43.68	6.74	47.82	10.01	45.96	8.31	47.62	9.93	46.66	8.29
Test 1992	61.73	10.09	59.08	9.39	61.95	10.19	58.86	9.02	61.78	10.12	58.98	9.25	61.61	10.02	58.36	9.51
Test 1994	62.70	9.13	61.42	8.09	62.85	9.21	61.16	7.89	62.54	9.20	62.18	7.88	62.93	8.93	59.07	8.42
Test 1996	65.16	9.90	67.70	7.50	65.13	10.07	67.35	7.05	65.00	9.97	68.32	6.99	65.16	9.86	68.80	6.39
Vulnerability	56.02	26.04	58.93	20.42	72.62	26.39	74.32	20.74	70.10	21.04	69.20	18.27	69.97	20.94	69.98	17.65
Total Previous Years	0.00	0.00	0.00	0.00	0.10	0.43	1.49	0.65	0.40	1.02	2.29	1.59	0.79	1.58	2.88	1.88
P900 Participation																
1990	0.00	0.00	1.00	0.00	0.04	0.20	0.58	0.49	0.09	0.28	0.37	0.48	0.11	0.32	0.26	0.44
1991	0.06	0.24	1.00	0.00	0.04	0.20	0.91	0.28	0.12	0.32	0.56	0.50	0.16	0.37	0.40	0.49
1992	0.08	0.27	0.73	0.44	0.00	0.00	1.00	0.00	0.09	0.28	0.61	0.49	0.14	0.34	0.41	0.49
1993	0.09	0.28	0.59	0.49	0.03	0.16	0.80	0.40	0.05	0.22	0.75	0.43	0.12	0.33	0.45	0.50
1994	0.10	0.30	0.41	0.49	0.07	0.25	0.54	0.50	0.00	0.00	1.00	0.00	0.10	0.30	0.59	0.49
1995	0.10	0.30	0.26	0.44	0.08	0.27	0.34	0.48	0.03	0.16	0.70	0.46	0.06	0.24	0.77	0.42
1996	0.07	0.26	0.18	0.39	0.06	0.24	0.23	0.42	0.04	0.20	0.37	0.48	0.00	0.00	1.00	0.00
Regional Dummies																
1st region	0.02	0.13	0.02	0.14	0.02	0.13	0.03	0.16	0.02	0.13	0.02	0.15	0.02	0.13	0.01	0.12
2nd region	0.02	0.13	0.01	0.09	0.02	0.13	0.01	0.10	0.02	0.13	0.01	0.10	0.02	0.13	0.02	0.13
3rd region	0.02	0.13	0.01	0.11	0.02	0.14	0.01	0.09	0.02	0.13	0.01	0.11	0.02	0.13	0.01	0.09
4th region	0.07	0.25	0.08	0.27	0.07	0.25	0.07	0.26	0.07	0.25	0.06	0.24	0.07	0.25	0.04	0.21
5th region	0.10	0.29	0.09	0.29	0.10	0.30	0.09	0.28	0.10	0.29	0.09	0.29	0.10	0.30	0.06	0.24
6th region	0.09	0.29	0.07	0.26	0.09	0.29	0.07	0.26	0.09	0.29	0.08	0.27	0.09	0.28	0.09	0.29
7th region	0.13	0.34	0.09	0.29	0.13	0.34	0.10	0.30	0.13	0.33	0.10	0.31	0.13	0.33	0.12	0.32
8th region	0.17	0.38	0.17	0.38	0.17	0.38	0.17	0.37	0.17	0.38	0.17	0.38	0.17	0.37	0.22	0.41
9th region	0.07	0.26	0.14	0.34	0.07	0.26	0.13	0.34	0.08	0.26	0.12	0.33	0.08	0.27	0.10	0.30
10th region	0.14	0.34	0.17	0.38	0.14	0.35	0.15	0.36	0.14	0.35	0.15	0.36	0.14	0.34	0.18	0.39
11th region	0.01	0.09	0.01	0.11	0.01	0.09	0.01	0.10	0.01	0.10	0.00	0.04	0.01	0.09	0.01	0.09
12th region	0.01	0.09	0.00	0.06	0.01	0.09	0.01	0.09	0.01	0.09	0.01	0.10	0.01	0.09	0.01	0.07

Table 2
Probit Regressions

	1992		1994		1996	
	Coef.	Std. Error	Coef.	Std. Error	Coef.	Std. Error
Previous Test	-0.0849	0.0077 *	-0.0224	0.0166 *	-0.0513	0.0092 *
Previous Vulnerability	-0.0095	0.0048 *	0.0164	0.0046 *	0.0136	0.0064 **
Sample Size	4149		4147		4149	
Probability of Predicting Correctly	69.30%		62.90%		67.30%	

Note: All regressions include regional dummies as well.
Significance at 5% is coded **, at 1% is coded *.

Table 3
Cross-Section OLS Regressions of Test on current P900 Status

	Coef	Std. Error
1992	-2.09	0.55 *
1994	1.31	0.42 *
1996	3.83	0.44 *

Cross Section Including Lagged P900 Status

	1992		1994		1996	
	Coef	Std. Error	Coef	Std. Error	Coef	Std. Error
t	-2.09	0.66 *	1.28	0.46 *	4.00	0.48 *
t-2	0.00	1.44	-2.08	1.18 **	-0.21	0.85
t-4			-0.23	0.91	-0.60	0.89
t-6					1.12	0.71

Note: all regression control for region, total previous participation
current and lagged vulnerability and lagged test score.

Table 4
Unrestricted GLS

	Test 1992		Test 1994		Test 1996	
	coef	std. Error	coef	std. Error	coef	std. Error
P900 1992	-1.81	0.61 *	-1.34	0.90	-1.70	0.75 **
P900 1994	-0.67	0.49	2.53	0.47 *	-1.95	0.62 *
P900 1996	-1.98	0.59 *	-3.42	0.47 *	4.26	0.47 *

Note: All equations have total_t, ivet, ive t-1, lagged test score and regional dummies with free parameters.
Significance at 10% is coded as ***, at 5% ** and at 1% *.

Table 5

Coef	Std. Error	Coef	Std. Error	coef	Std. Error
I. Presence of School Effect Model: Non linear restriction imposed					
b**1**		b**2**		b**3**	
-1.12	0.76	3.90	1.03 *	9.34	3.37 *
l**1**		l**2**		l**3**	
-0.67	0.34 **	-0.74	0.39 ***	-1.92	0.53 *
g**1**		g**2**		g**3**	
		1.80	0.59 *	2.64	1.56 ***
c**2 (1)**	**0.06**				
II. Restrict P900 coefficient to be constant: one program effect					
b					
3.21	0.39 *				
l**1**		l**2**		l**3**	
-4.10	0.57 *	-0.87	0.39 **	-2.73	0.46 *
		g**2**		g**3**	
		0.84	0.16 *	-0.05	0.15
c**2 (3)**	**21.83**				
III. Restrict fixed school effect: g**=1**					
b**1**		b**2**		b**3**	
-0.18	0.77	3.32	0.67 *	6.60	0.62 *
l**1**		l**2**		l**3**	
-1.33	0.47 *	-1.01	0.38 *	-2.66	0.34 *
c**2 (3)**	**6.27**				
IV. Restrict constant correlation coefficient: equal lambdas					
b**1**		b**2**		b**3**	
-0.26	0.75	5.34	0.68 *	5.69	0.69 *
l					
-1.27	0.32 *				
		g**2**		g**3**	
		2.31	0.69 *	1.29	0.52 **
c**2 (3)**	**7.11**				

V. Restricted constant correlation coefficient and school fixed effects=1					
b**1**		b**2**		b**3**	
0.63	0.65	4.22	0.56 *	5.58	0.51 *
l					
-1.88	0.22 *				
c**2 (5)**	**16.31**				
VI. Fixed Effects Model					
b					
3.79	0.42 *				
l**1**		l**2**		l**3**	
-3.14	0.38 *	-1.64	0.31 *	-1.60	0.29 *
c**2 (5)**	**49.99**				
VII. Total Fixed Effects Model					
b					
3.00	0.33 *				
l**1**		l**2**		l**3**	
-1.25	0.01 *	-0.56	0.33 ***	-2.24	0.21 *
c**2 (37)**	**14062.78**				

References

Carlson, B. (2000). Achieving Educational Quality: What Schools Teach Us. Learning from Chile's P900 Primary Schools. *Desarrollo Productivo Series*, 64, ECLAC.

Chamberlain, G. (1982). Multivariate Regression Models for Panel Data*, Journal of Economics*, 18, 5-46.

Cox, C. (1997). La Reforma de la Educación Chilena: Contexto, Contenidos, Implementación, Programa de Promoción de la Reforma Educativa en América Latina (PREAL).

Espinola, V. (1993). *The Educational Reform of the Military Regime in Chile: The System's Response to Competition, Choice, and Market Relations.* Unpublished Ph.D. dissertation, University of Wales, United Kingdom.

Friedman, M. (1955). The Role of Government in Education. In R. A. Solo (Ed.), *Economics and the Public Interest* (pp. 123-144). New Brunswick, NJ: Rutgers University Press.

Greene, W. H. (1997). *Econometric Analysis*. (3rd ed.). Upper Saddle River, NJ: Prentice Hall.

Goldberger, A. (1981). Linear Regression after Selection, *Journal of Econometrics*, 15, 1981.

Hausman, J. (1978). Specification Tests in Econometrics, *Econometrica,* 46, 1251-1272.

Heckman, J. J. (1979). Sample Selection Bias as a Specification Error. *Econometrica, 47*(1), 153-161.

Jakubson, G. (1991). Estimation and Testing of the Union Wage Effect Using Panel Data, *The Review of Economic Studies*, Volume 58, Issue 5, Oct. 1991, 971-991.

Lee, L.-F. (1983). Generalized Econometric Models with Selectivity. *Econometrica, 51*(2), 507-512.

MINEDUC (2000). Evaluación del Programa de las 900 Escuelas. División de Educación General, Chilean Ministry of Education, Dec. 2000.

Footnotes

---

[i] Since 1981 there are three types of schools in Chile: Private paid, public subsidized and private subsidized. Around 90% or 9000 of the schools are subsidized and therefore the 900 schools included in the Program, which are only private subsidized and public schools, represent around 10% of those schools.

[ii] Schools in Chile take a standardized test at 4^{th} grade every even year and 8^{th} grade every odd year, starting in 1988. The test scores reported are school averages. Given the time pattern of the tests, students are never tested more than once, either at 4^{th} or 8^{th} grade, and therefore any value added analysis is obscured by the fact the students are not the same (cohort effect).

[iii] This section follows closely Jakubson's (1991) paper.

[iv] C_s may also be correlated with the other X's, but since we are only interested in estimating an unbiased treatment effect we will assume for simplicity that it is only correlated with the P900 dummy.

[v] The economic vulnerability index is coded from 0 to 100, being 100 the most vulnerable. Its definition changes yearly in terms of the characteristics included and the way they are weighted (i.e. it is not strictly comparable between years), but always tries to capture the economic background of the students. The elementary school index is used whenever it is available, otherwise the aggregate index is used.

[vi] This would be the case if the administration became more efficient in delivering the resources or if the training focused in short run effects.

[vii] Variables are included as deviations from the mean.

[viii] Corresponds to imposing the restrictions in (5) above.

[ix] γ_1 is normalized to 1.

Documentos de Trabajo
Banco Central de Chile

NÚMEROS ANTERIORES

Working Papers
Central Bank of Chile

PAST ISSUES

La serie de Documentos de Trabajo en versión PDF puede obtenerse gratis en la dirección electrónica: http://www.bcentral.cl/Estudios/DTBC/doctrab.htm. Existe la posibilidad de solicitar una copia impresa con un costo de $500 si es dentro de Chile y US$12 si es para fuera de Chile. Las solicitudes se pueden hacer por fax: (56-2) 6702231 o a través de correo electrónico: bcch@condor.bcentral.cl

Working Papers in PDF format can be downloaded free of charge from: http://www.bcentral.cl/Estudios/DTBC/doctrab.htm. Printed versions can be ordered individually for US$12 per copy (for orders inside Chile the charge is Ch$500.) Orders can be placed by fax: (56-2) 6702231 or e-mail: bcch@condor.bcentral.cl

DTBC-169 Julio 2002
Industrial Policies and Growth: Lessons from International Experience
Marcus Noland y Howard Pack

DTBC-168 Julio 2002
Quantity and Quality of Economic Growth
Robert J. Barro

DTBC-167 Julio 2002
Monetary Union: European Lessons, Latin American Prospects
Eduard Hochreiter, Klaus Schmidt-Hebbel y Georg Winckler

DTBC-166 Junio 2002
Monetary Policy Implementation and Results in Twenty Inflation-Targeting Countries
Klaus Schmidt-Hebbel y Matías Tapia

DTBC-165 Junio 2002
Estimating Gaps and Trends for the Chilean Economy
Gabriela Contreras M. y Pablo García S.

DTBC-164 Junio 2002
It's Not Factor Accumulation: Stylized Facts and Growth Models
William Easterly y Ross Levine

DTBC-163 Junio 2002
Macroeconomic Management in Emerging Economies and the International Financial Architecture
José De Gregorio

DTBC-162 Junio 2002
Two-Part Tariff Competition with Switching Costs and Sales Agents
Solange Berstein

DTBC-161 Junio 2002
Saving and Life Insurance Holdings at Boston University – A Unique Case Study
B. D. Bernheim, Solange Berstein, Jagadeesh Gokhale y L. J. Kotlikoff

DTBC-160 Junio 2002
The Federal Design of a Central Bank in a Monetary Union: The Case of the European System of Central Banks
Sylvester C.W. Eijffinger

DTBC-159 Junio 2002
Testing Real Business Cycles Models in an Emerging Economy
Raphael Bergoeing y Raimundo Soto

DTBC-158 Junio 2002
Funciones Agregadas de Inversión para la Economía Chilena
Héctor Felipe Bravo y Jorge Enrique Restrepo

DTBC-157 Mayo 2002
Finance and Growth: New Evidence and Policy Analyses for Chile
Ross Levine y María Carkovic

DTBC-156 Mayo 2002
The Effects of Business Cycles on Growth
Antonio Fatás

DTBC-155 Mayo 2002
Trends, Cycles and Convergence
Andrew Harvey

DTBC-154 Mayo 2002
Coping with Chile's External Vulnerability: A Financial Problem
Ricardo J. Caballero