Poor students have it tougher: academic performance by ninth-graders in two Mexican states

Hector J. Villarreal*

EGAP/ITESM Campus Monterrey

(this version 05/2007)

Abstract This paper explores the correlations that exist between standardized test scores in seven subjects and sociodemographic variables for ninth-graders in two Mexican states. The analysis is performed with two basic sets of variables: identity or circumstances, which cannot change easily such as language spoken at home, and behavior variables like working outside the household. Additionally, we control the specific school attended, and participation in a CCT (Oportunidades). Results show, that some identity variables present high positive correlations with the test scores, especially maternal education. The latter suggests a disadvantage for poorer students which cannot be easily surpassed with effort/behavior. While the majority of the schools do not show significant correlations with the scores, there were some "star" and "failure" schools in the sample. Trying to get the less favored students in the first group may be a sensible policy. Finally, once other variables are taken into consideration, participation in the CCT has a low correlation with the scores.

Keywords: Standardized tests, inequality, education, México.

JEL: D31, H52, I21, I32.

Contact Information: hjvp@itesm.mx

*I am very grateful to my colleague Guadalupe Villarreal for the invitation to work with the dataset employed in this paper

1 Introduction

The last ten years have witnessed the resurgence of important research efforts to understand inequality. This coincides with a period of rapid economic expansion in the world and the widening of inequality between and within countries, (Albuquerque 2004). Meanwhile, even with differences in degree, equality remains an important value for most societies. However, there has been an important change in public policy: *equality of outcome* is still sought, but much more emphasis is placed in *equality of opportunity*. The later idea referred many times as leveling the (starting) field, implies that inequality is more tolerable if it is caused by merit or effort than by passed disadvantages. A broad discussion can be found in (Arrow, Bowles, and Durlauf 2000).

Related to the idea of equality of opportunity is the notion that many households are *trapped in poverty:* low-income households invest very little on education and health, perpetuating poverty via vicious circles. Popular instrument nowadays to break poverty traps are *conditional cash transfers* programs (CCT), (Lindert, Skoufias, and Shapiro 2005; Das, Do, and Özler 2004, 2005; Bourguignon, Ferreira, and Leite 2003; Rawlings and Rubio 2005). These programs possess two peculiarities described in their name. On the one hand, CCT involve direct cash transfers as opposed to "in kind" transfers or to subsidies of particular goods. In theory, the "cash transfer" element is associated with short-run welfare alleviation. The second characteristic is that reception of resources implies the commitment of the household to follow certain behavior (school attendance by children, medical check-ups, etc.). The objective within this second element is to align incentives in order for poor households to accumulate human capital. This second element is expected to enhance welfare in the longer run. Consequently CCT should improve equality of outcomes and opportunities.

A sensible question from a policy perspective is how the accumulation of human capital of participants in CCT varies in comparison to the accumulation of non-participants. With respect to years of schooling several assessments exist, (for example Schultz 2000; Skoufias and Parker 2001; Behrman, Parker, and Todd 2005). Besides the quantity component (years of schooling), a less explored quality component of human capital can be of significant importance (Hernandez-Zavala, Patrinos, Sakeallariou, and Shapiro 2006). While it is difficult to observe the quality of human capital, scores in standardized tests may be a sensible proxy for the quality of the educational component of human capital (Garza and Villarreal 2007). A model presented in

that paper predicts that participants in CCT should have on average lower grades. The result is caused by two effects: first, a selection effect (*switching*) where some students may decide to enroll because of the transfer and will minimize effort. Second, the effects on grades of variables that affect (or are affected by) the income of the family and are not equalized after the cash transfer (*incomplete homologation*). Along a similar research line, (Bourguignon, Ferreira, and Menéndez 2005) explore a decomposition of inequality in earnings at Brazil due to circumstances, which lie beyond the control of the individual; and effort variables, controlled by the person.[1]

In this paper we investigate the incomplete homologation (circumstance) effect. The data is presented in the next section. We have a sample of 1243 ninth grade students in Mexico and their exam results at several subjects. A subset of these students participates in a famous Mexican CCT, i.e. Progresa/Oportunidades. The availability of a rich set of sociodemographic and personal variables permits the testing of alternative econometric specifications. Descriptive statistics show that participants in the CCT do have lower scores, and these are clustered by school. The latter could suggest that a variance decomposition should be performed in order to understand how much of the difference is attributed to the school. However, given that many of the explanatory variables present endogenous relations, a simple decomposition may be misleading. The econometric specifications and statistical tests are included in the third section. The final section briefly presents the conclusions of the analysis and suggests some policy recommendations.

2 The Data

The dataset includes 1243 students that finished middle school in 2005, and is drawn from a survey conducted in April of the same year by Dr. Guadalupe Villarreal. The dataset contains information about students from two states, Chiapas and Nuevo León, studying in two types of schools, telesecondaries and general secondary schools. The students in the survey took the EXANI-I test (discussed below) and filled a registry form that was used to obtain sociodemographic data.

Table 1 provides an outlook of Chiapas and Nuevo León, swhowing the contrast in development levels in the two Mexican states. On the one hand, Nuevo León is a state in the Northeastern part of the country (it has a small

[1] This classification is due to (Roemer 1998).

border with United States), with most of its population living in a large metropolitan area. The average schooling of its adult population is slightly above nine. Chiapas is a state in the southeastern part of the country. More than half of its population lives in rural areas, including a considerable percentage of indigenous people. The average schooling of its adult population is around six. The differences between the two states should prove very useful in the identification of variables affecting the education experience.

The data was obtained by a two stage stratified sampling procedure. In the first stage schools were classified in homogenous strata according to four variables:

a) State, (Chiapas or Nuevo León)

b) Type of School, (General or Telesecondary)

c) The deprivation level of the area where the school is located [Source: CONAPO], (very low, low, medium, high, or very high)[2]

d) If the school is compensated by any program of Consejo Nacional de Fomento Educativo [CONAFE], (yes or no).[3]

Table 1. Basic Information About Chiapas and Nuevo León

	Chiapas	Nuevo León
Total Population	3,920,892	3,834,141
Urban Population	46%	93%
Literacy	77%	95.6%
Number of Middle Schools	1,403	12,991
Number of Households	778,845	878,600
Houses with sewerage	74.4%	95.5%
Houses with Electricity	87.9%	98.5%
Houses with sewer serv.	62.3%	90.8%
Families in Oportunidades	554,525	49,564
Oportundiades Scholarships*	153,326	6,864
Gini Coefficient**	0.5729	0.5185
Human Development Index‡	0.6926	0.8425
Deprivation Index┳	2.2507	-1.3926

Source: INEGI, Anuario Estadístico por Entidad Federativa; available from wwww.inegi.gob.mx

[2]Very low is the least level of deprivation.

[3]These are social programs aimed to improve the supply side of education targeted at low performance schools.

*Only for secondary school
**Authors' calculations, source: ENIGH 2004, INEGI
‡Source: CONAPO, Índice de Desarrollo Humano 2000; avaialble from www.conapo.gob.mx
ŦSource: CONAPO, Índice de Marginación 2000; available from www.conapo.gob.mx

In the second stage weighted random sampling was used within each stratum. The observed unit was the school, and its probability of being sampled was proportional to the number of registered students it had in ninth grade. EXANI was designed to have a population mean of 1,000 points, and a standard deviation of 100 points.The sample for each stratum complies with an error (with respect to the population mean of the respective stratum) of two EXANI points in 95% of the cases.

Tables 2 and 3 describe the samples for Chiapas and Nuevo León respectively. The include the type of school (type), the deprivation level (Deprivation), if the school is compensated (Comp.), the number of schools sampled in each stratum (# of Schools), students not participating in oportunidades and students participating (Stu. wo/Oport., Stu w/Oport), plus the total number of students sampled in each stratum.

Table 2 Chiapas Sample

Type	Deprivation	Comp.	# of Schools	Stu. wo/Oport.	Stu w/Oport.	Total
Gen.	very low	no	1	29	70	99
Gen.	low	no	1	5	32	37
Gen.	very high	no	1	3	67	70
Tel.	very low	yes	1	11	14	25
Tel.	very low	no	2	13	68	81
Tel.	low	yes	2	15	59	74
Tel.	low	no	1	0	14	14
Tel.	medium	yes	3	12	103	115
Tel.	medium	no	1	1	35	36
Tel.	high	yes	1	1	20	21
Tel.	high	no	2	10	69	79
Tel.	very high	yes	2	10	33	43
			18	110	584	694

Table 3 Nuevo León Sample						
Type	Marg.	Comp.	# of Schools	Stu. wo/Oport.	Stu w/Oport.	Total
Gen.	very low	no	2	202	28	230
Gen.	low	no	2	71	53	124
Gen.	medium	no	1	59	11	70
Gen.	high	no	1	30	40	70
Tel.	medium	yes	2	2	12	14
Tel.	high	yes	3	11	33	44
Tel.	very high	yes	2	5	25	30
			12	380	202	582

EXANI is a complex standardized test. It was developed by the Mexican government to assess both abilities (verbal and quantitative), and command of different subjects (math, Spanish, physics, chemistry, biology, history, geography and citizenship behavior[4]). Table 4 provides the average test scores by subjects for our sample according to state and participation in the program.

Table 4 Average EXANI Scores				
	Chiapas		Nuevo León	
	Oport.	N.Prog.	Oport.	N.Prog.
Math	867	883	905	938
Spanish	855	904	912	968
Physics	856	882	898	928
Chemistry	858	885	888	898
Biology	872	886	899	930
Geography	864	884	877	921
History	870	882	899	930

3 Econometric Analysis

The econometric analysis was performed using the test scores as dependent variables. Given the large amount of sociodemographic variables provided

[4]Citizenship behavior will not be considered in this paper. For additional information about the EXANI-I, the reader can visit the CENEVAL's webpage (http://www.ceneval.edu.mx/portalceneval/index.php?q=info.fichas.ficha1)

in the EXANI questionnaire it was important to select which ones should be used as independent variables. Since many of the variables would be highly correlated, simple variance decompositions may be misleading. Instead, great emphasis was placed in identification issues, that is, predicted effects (correlations) that could be backed up with theory. Given the incomplete homologation hypothesis to be tested, a first group of variables that would not change (at least easily) due to the program was selected, they are catalogued as circumstance or identity. These include: the parents education (both father and mother in school years), the number of siblings, if whether Spanish is the language spoken at home, whether the household is in a rural community, and the state where the household is located (i.e. Chiapas or Nuevo León). If the student attends a telesecondary[5], it reflects the living standards of the community so it can be catalogued within the first group of variables, but it also has to do with the school experience. A second group of variables are believed to be behavior characteristics: working, that describes if the student works outside the household; study hours, the number of hours the student devotes to study; and egg servings a week[6]. The variable fast learning is a dummy taking the value of one if the student declares herself as a fast learner or moderately fast learner, zero otherwise. Finally, the variable income refers to a self-chosen household income level from a set of brackets. It is possible that it has considerable measurement error, thus it will be used primarily for identification in a second stage econometric analysis.

It is interesting to note the huge differences in the levels of almost all the independent variables between Chiapas' households participating in Oportunidades and Nuevo León's households that do not participate in the program. However, the levels of the variables for non participating households in Chiapas with participating households in Nuevo León result much more similar. This coincides with the results from the test scores described in Table 4. That both patterns fit together may suggest that the explanatory variables chosen will aid in the identification of correlations between sociodemographic

[5]Telesecondary is a distance education format available in rural communities in which the student receives the classes through television. This system has come to alleviate the problem of the shortage of middle school institutions and teachers in isolated communities. The general middle school is the traditional format institution where the teacher gives the classes directly to the students.

[6]This variable was chosen as a proxy of nutrition. The main reason is that eggs provide high quality protein at a relative low price. Alternatively a basket could have been chosen but would require extra assumptions.

characteristics and test scores.

Table 5 Independent Variables				
	Chiapas		Nuevo León	
	Oport.	N.Prog.	Oport.	N.Prog.
father's education	5.0	6.9	6.3	9.5
mother's education	4.2	5.9	6.3	9.5
# of siblings	4.5	3.8	3.7	2.4
income	2249	2279	2757	4026
Spanish	0.67	0.84	0.99	1.00
rural	0.88	0.74	0.74	0.52
telesecondary	0.71	0.66	0.34	0.04
working	0.14	0.06	0.10	0.04
study hours	3.3	3.8	3.7	4.5
egg servings	1.9	2.4	3.2	3.5
fast learning	0.70	0.80	0.79	0.77

Our first stage analysis explores the correlations between test scores, the two sets of variables described above, plus constructed dummies representing each of the schools in the sample (as shown in tables 2 and 3, the sample includes 30 different schools). The specification will be the following:

$Y_i = \beta X_i + \eta S_i + e_i$

where y_i are the test scores, X_i are the identity and behavior independent variables, S_i the dummies representing each school, β and η are conformable vector of parameters and e_i an error term. Each equation will be estimated independently using OLS. Alternatively the system of seven equations could have been SUR estimated (Zellner 1962). We selected the former because emphasis will be placed in individual variable tests rather than parameter significance tests across equations.

Given that a sample of thirty schools would produce a very long list of independent variables preliminary econometric tests were performed. The criteria followed was that schools would be kept in the list if they were significantly correlated (at a 10% level) with at least one of the test scores. Six schools out of the total thirty were selected (implications will be discussed in he final section of the paper). Four are in Chiapas: s6, normal, not compensated, rural; s12, normal, not compensated, rural; s16, normal, compensated, rural; and s17, telesecondary, not compensated, rural. The other

two are in Nuevo Leon: s28, telesecondary, not compensated, rural; and s30, telesecondary, not compensated, rural[7].

Tables 6a-6c summarize the results. Table 6a concentrates the effects we labeled "identity" (circumstantial). With respect to the parents education, the fathers' education appears to have a lesser role, not economically or statistically significant for any test with the exception of physics. On the other hand, the mothers' education seems to have a critical role, it is very significant with the exception of physics (a possible complementarity). After controlling for variables in Table 5, the mothers' education can explain a considerable amount of the students' test scores. Other identity variables present markedly smaller correlations with the test scores. Living in a rural community was not significant in any of the equations. The dummy variable that controls for state (Chiapas or Nuevo León) was only statistically significant in two out of the seven equations, and in those the effects have opposite signs (being Nuevo León correlated positively with Spanish and negatively with geography). This result is not trivial. From the states' characteristics described in Table 1 a sensible assumption would be that both states had very different educational systems, which would explain the existent differences in test scores. However, once household characteristics are controlled (and specific schools as will be explained below), the correlation is highly diminished. A huge negative effect was expected in the telesecondary systems, it only shows significant in the math test score. Similarly, number of siblings was only important (in this case with a negative sign) on the chemistry test score.[8]Being Spanish the language spoken at home, and participating in Oportunidades do show economic and statistically significant effects. While the math score is not significantly affected, Spanish is strongly positively correlated with the physics, chemistry, biology and history scores, while Oportunidades is negatively correlated with the Spanish and geography scores. It is very possible that both variables are capturing the same effect, an issue that will be discussed in the next section.

The effects of the "behavior" variables are summarized in table 6b. As expected, working has a negative significant effect in some of the test scores (Spanish, chemistry and biology). Not surprisingly, studied hours is very significant both statistically and economically in all the scores. Interestingly,

[7]The specific characteristics of the other twenty-four can be provided upon request.

[8]While some conjectures emerge from the math/telesecondary relation, i.e. teachers not supervising students' practice. We do not have any hypothesis supporting the chemistry/siblings correlation.

while there is a negative correlation between working and studied hours (students who work outsides their homes report studying less hours), the magnitude is very small. The amount of egg portions has a small but non negligible effect on three test scores (math, physics and geography), suggesting that nutrition does play a role[9]. Finally, the self-defined fast learning ability is significant in almost all the test scores with the exception of biology. The result is important, however future research is needed to highlight if the variable captures true ability, motivation, or other characteristic.

Table 6c describes the school effect. Within our sample of thirty schools, twenty-four seem not to be explaining the differences in scores between the 1,243 students of the sample. This should be interpreted in the following way: if a student has average values for all the other explanatory variables and goes to one of the twenty-four mentioned schools, since the model specification is linear her expected test scores are the expected values of the sample. On the other hand, students going to the other six schools of the sample could have expected test scores different from the sample means. From the six schools, four are in Chiapas. These four schools "push" students' scores downward, with school #6 in the sample having particularly large negative effects followed by school #16 in the sample. The two other schools are in Nuevo León, and "pull" students' scores upward. Interestingly, school #28 in the sample has a very large positive effect in the math test score despite of being a telesecondary.

3.1 Endogeneity

Of the independent variables employed in the previous set of regressions we are especially interested in the effects that participating in Oportunidades has on the test scores. However, as discussed in the beginning of the paper it would typically be correlated with other variables that may influence the test scores. Thus it is difficult to isolate the effects, an econometric problem of endogeneity. So, in order to test the robustness of our first stage econometric analysis we will use an alternative econometric specification. Following Maddala (1983), participation in Oportunidades will be modeled as a treatment effect:

$Y_i = \beta X_i + \delta Z_i + e_i$

In this case Z_i is an endogenous dummy variable indicating whether the

[9]For a detailed discussion see (Paxson and Schady 2005).

treatment is assigned (participation in Oportunidades). The happening of the treatment is modeled as the outcome of an unobserved latent variable Z_i^*. We will assume Z_i^* is a linear function of the exogenous covariates W_i and a random component u_i:

$Z_i^* = \gamma W_i + u_i$

with the observed treatment:

$$Z_i^* = \left\{ \begin{array}{c} 1, \text{if } Z_i^* > 0 \\ 0, \text{otherwise} \end{array} \right\}$$

It is assumed that e and u are bivariate normal with mean zero and covariance matrix:

$$\left[\begin{array}{cc} \sigma & \rho \\ \rho & 1 \end{array} \right]$$

which completes the specification. Some details about interpretation and the estimation method are presented in Appendix A.

Tables 7a-7c summarize the results of this section. Table 7a considers the identity variables. The results for parents' education remains very important and qualitatively similar to the previous results, the only considerable difference is that the mother's education is not significant in explaining the geography score (compared to the previous specification). With respect to the other variables in this set: "state" remains significant with a negative sign for geography but not Spanish. Telesecondary retains a significant negative correlation with respect to the math score and now also with respect to physics, although with smaller magnitude in the latter. The number of siblings is negatively correlated with the chemistry score. Spanish spoken at home continues to be correlated positively and significantly with the physics, chemistry, biology and history scores; however now the set is augmented with the math score whose correlation was not significant in the previous specification.[10]Now, as explained previously the effect of participating in Oportunidades is of special interest, once the treatment effect is considered, the correlation of the program with all the scores (except geography) is not significant. The case of the geography score is very interesting, it seems that the OLS specification (as discussed in the appendix) was underestimating the effect (in this case negative) of participating in the program. While the

[10]The Spanish score does not shows a statistically signifcant correlation with Spanish being spoken at home.

correlation was already negative and statistically significant in the previous section, its economic significance increases substantially in the second specification. A small discussion of this finding is provided in the last section.

The effects of "behavior" variables are shown in table 7b. Similarly to the previous section, working outside has a negative effect, and is significant in three of the scores. Nonetheless, the most economic significant effect comes from studied hours, being positive and statistically significant for every test score. The effects for egg portions and the self-defined fast learning are almost identical to the previous section. The results for the "specific schools", shown in table 7c, are also very similar to the ones found in the OLS analysis of the previous section.

4 Conclusions

This paper explores the correlation that exists between standardized test scores in seven subjects and sociodemographic variables for ninth-graders in two Mexican states. A considerable portion of the students in the sample participate in a famous Mexican CCT (Oportunidades). Scores in standardized tests may be used as a proxy for the quality in schooling (as opposed to years of school that is typically associated with the quantity). Furthermore, the scores may predict both future human capital accumulation and performance in the labor market. Four characteristics of the students are hypothesized to affect their school experience: identity or circumstantial variables, behavior variables, the specific school attended and participation in the program. We believe that identified effects in each of these groups of variables may require different support policies.

One of the findings of this study is that "state" (living in Chiapas or Nuevo León) is not significant in explaining the test scores once other variables are considered. The same occurred for urban versus rural populations (no statistical difference). Both results came as a surprise given the different development levels they present and would defy simple descriptive statistics. On the other hand the mother's education and Spanish spoken at home do play a critical role in explaining the differences in scores. These results are aligned with the findings of (Bourguignon, Ferreira, and Menéndez 2005). Given the high correlation of poverty with low maternal education and indigenous background, the results suggest a handicap for poorer students.

With respect to behavior variables, the amount of weekly studied hours correlates positively and highly significantly with all the test scores. This was expected, its caveat is that students in the sample study outside school (including homework time) a ludicrous average of four hours per week. Self-defined "fast-learning" ability was significant in explaining the test scores but to a lesser degree than studied hours. Egg portions per week were significant in some of the test scores suggesting the importance of nutrition.

Another interesting finding is that while attending school obviously matter, most of the schools in our sample (80%) do not appear to have a major effect deviating students from the test scores sample mean. One fifth (20%) of the sampled schools do show significant correlations with the test scores. Of these, some of them seem to have a very negative effect on scores (four schools in the sample, all of them in Chiapas), and some to have a very positive effect on scores (two schools in the sample, both in Nuevo León). While it is tempting to suggest that rich states are more probable to provide "star" public schools and the opposite for poorer states, the sample in this paper does not allow doing that generalization. It remains a question for further research.

Finally, with respect to the very unexplored question of effects of CCTs on standardized test scores,[11]we did not find strong evidence of signficant correlations, with the exception of one subject[12]. The first econometric specification relied upon OLS, (here there was a significant correlation in two subjects). However, given the endogenous relations of many of the variables, a second specification was employed that controlled participation in the program via a treatment effect, results suggest that OLS analysis overestimates the effects of participating in the program on scores. A possible reconciliation of the low correlations between participation in the CCT and test scores, with observed descriptive statistics is the following: it is not that participation in the program lowers the scores, but participants have on average characteristics that do correlate negatively (e.g. lower maternal education, etc.).

Under the assumption that standardized tests reflect the quality of the

[11]The analysis performed in this paper should not be regarded as an evaluation of Oportunidades in any sense. The intention of the paper is to explore the correlation between some sociodemographic variables of poor students and their grades in standardized test scores.

[12]A limitation of this paper is that we are not exploring selection bias, which may play a role. For a discussion refer to (Garza and Villarreal 2007).

acquired human capital (educational element of it) this paper suggests that poorer students have an inherent disadvantage. That is, the attendance to a similar number of school years compared to non-poor students does not guarantee an automatic leveling of the field. The rationale behind the latter is that "identity" or circumstantial variables like their parents education, the language spoken at home, etc., put an extra burden on education. To the extent that governments want to compensate less favored students, several policies may help, for example: a heavier exposure to the national language, study groups after school (with the adequate incentives), etc. With respect to schools, if poorer students do not have access to the "star" public schools of the system, it is important to control that they do not get into the worst.

References

[1] Albuquerque, Pedro (2003). A practical log-linear aggregation method with examples: heterogeneous income growth in the USA. Journal of Applied Econometrics, 18, 665-678.

[2] Arrow, Kenneth, Samuel Bowles, and Steven Durlauf, eds. (2000). Meritocracy and Economic Inequality.Princeton: Princeton University Press.

[3] Behrman, Jere R., Susan W. Parker, and Petra E. Todd (2005) Long-Term Impacts of the Oportunidades Conditional Cash Transfer Program on Rural Youth in Mexico. Ibero America Institute for Econ. Research (IAI) Discussion Papers 122, Ibero-America Institute for Economic Research.

[4] Bourguignon, François, Francisco H. G. Ferreria, and Phillippe G. Leite (2003). Conditional Cash Transfers, Schooling, and Child Labor: Micro-Simulating Brazil's Bolsa Esola Program. The World Bank Economic Review, 17, 229-

[5] Bourguignon, François, Francisco H. G. Ferreria, and Marta Menéndez (2005). Inequality of Opportunity in Brazil. World Bank Research Working Paper.

[6] Das, Jishnu, Quy-Toan Do, and Berk Özler (2004). Conditional Cash Transfers and the Equity-Efficiency Debate. Policy Research Working Paper Series 3280, The World Bank..

[7] Das, Jishnu, Quy-Toan Do, and Berk Özler (2005). Reassessing Conditional Cash Transfer Programs. World Bank Research Observer, 20, 57-80.

[8] Garza, Mabel and Hector J. Villarreal (2007). Do Conditional Cash Transfers Affect Poor Students' Performance? Working Paper EGAP / ITESM Campus Monterrey.

[9] Hernadez-Zavala, Martha, Harry A. Patrinos, Chris Sakellariou, and Josehph Shapiro (2006). Quality of Schooling and Quality of Schools for Indigenous Students in Guatemala, Mexico and Peru. Policy Research Working Paper Series 3982, The World Bank.

[10] Lindert, Kathy, Emmanuel Skoufias, and Joseph Shapiro (2005) Redistribution of Income to the Poor and the Rich: Public Transfers in Latin America and the Caribbean. Working Paper The World Bank.

[11] Maddala, G.S. (1983). Limited-dependent and Qualitative Variables in Econometrics. Cambridge: Cambridge University Press.

[12] Paxson, Christina and Norbert Schady (2005). Cognitive development among young children in Ecuador : the roles of wealth, health and parenting. Policy Research Working Paper Series 3605, The World Bank.

[13] Rawlings, Laura B. and Gloria Rubio (2003). Evaluating the Impact of Conditional Cash Transfer Programs. Lessons from America Latina. Policy Research Working Paper Series 3119, The World Bank.

[14] Roemer, John E. (1998). Equality of Opportunity. Cambridge, MA: Harvard University Press.

[15] Schultz, T. Paul (2004). School Subsidies for the Poor: Evaluating the Mexican Progresa Poverty Program. Journal of Development Economics, 74, 199-250.

[16] Skoufias, Emmanuel and Susan W. Parker (2001). Conditional cash transfers and their impact on child work and schooling. FCND discussion papers 123, International Food Policy Research Institute (IFPRI).

[17] Zellner, Arnold (1962). An efficient method of estimating seeming unrelated regressions and tests for aggregation bias. Journal of the American Statistical Association, 57, 348-368.

Appendix A

Our estimation method for the model specification presented in the endogeneity section will be Maximum Likelihood. Maddala (1983) derives the likelihood function for this model:

$Y_i = \beta X_i + \delta Z_i + e_i$

$Z_i^* = \gamma W_i + u_i$

with the observed treatment:

$$Z_i^* = \left\{ \begin{array}{c} 1, \text{if } Z_i^* > 0 \\ 0, \text{otherwise} \end{array} \right\}$$

if e and u are bivariate normal with mean zero and covariance matrix:

$$\left[\begin{array}{cc} \sigma & \rho \\ \rho & 1 \end{array} \right]$$

then:

$$\ln L_i = \ln \Phi \left(\frac{\gamma W_i + (Y_i - \beta X_i - \delta)\rho/\sigma}{\sqrt{1-\rho^2}} \right) - \frac{1}{2} \left(\frac{Y_i - \beta X_i - \delta}{\sigma} \right)^2 - \ln(\sqrt{2\pi}\sigma), \text{ if } Z_i = 1$$

and

$$\ln L_i = \ln \Phi \left(\frac{-\gamma W_i - (Y_i - \beta X_i)\rho/\sigma}{\sqrt{1-\rho^2}} \right) - \frac{1}{2} \left(\frac{Y_i - \beta X_i}{\sigma} \right)^2 - \ln(\sqrt{2\pi}\sigma), \text{ if } Z_i = 0$$

where Φ is the standard normal cumulative density function.

Given this specification the difference in expected grades between participants and nonparticipants in Oportunidades is:

$$E(Y_i|Z_i = 1) - E(Y_i|Z_i = 0) = \delta + \rho\sigma \left[\frac{\phi(\gamma W_i)}{\Phi(\gamma W_i)\{1 - \Phi(\gamma W_i)\}} \right]$$

being ϕ the standard normal density.

For space reasons the results of the treatment equation are not listed (they could be provided upon request). The variables used to explain participation in Oportunidades (treatment), were all the behavior variables plus the self-reported income bracket. Results were very consistent across equations all the variables were highly significant with the exception of telesecondary (this

was expected), and Spanish spoken at home (this was not expected). Also the ρ in five out of seven equations was positive, suggesting that OLS was overestimating the effect of participating in the program.

Econometric Results

Table 6

	Math	Span.	Phys.	Chem.	Bio.	His.	Geo.
cons	830.7*	818.3*	822.3*	832.0*	825.9	813.7*	827.1*
s.e.	20.0	18.8	19.4	17.4	19.1	17.8	17.6
fath. edu	1.7	1.0	2.8*	0.1	0.6	1.4	2.0*
s-e	1.1	1.0	1.1	1.0	1.1	1.0	1.0
moth. edu	2.9*	4.5*	0.6	1.7**	3.7*	3.1*	2.1*
s.e.	1.1	1.1	1.1	1.0	1.1	1.0	1.0
state	13.9	18.5*	11.4	-1.3	-2.9	-3.2	-14.6**
s.e.	9.0	8.5	8.7	7.8	8.6	8.0	7.9
tele.	-41.9*	6.9	-15.6	7.2	-1.7	-3.0	4.8
s.e.	10.1	9.5	9.8	8.8	9.7	9.1	8.9
rural	-13.7	-11.9	-14.3	8.6	-6.5	-11.5	-6.2
s.e.	10.1	9.5	9.8	8.8	9.6	9.0	8.9
sibl.	-1.12	0.1	-2.2	-3.8*	-0.6	0.2	-0.9
s.e.	1.7	1.6	1.7	1.5	1.6	1.5	1.5
Spanish	17.1	13.9	19.1**	15.2**	29.6*	28.5*	11.8
s.e.	10.6	10.0	10.3	9.23	10.2	9.5	9.3
Oport.	-5.3	-23.7*	-3.5	-4.9	-4.7	-1.2	-12.1**
s.e.	7.8	7.3	7.5	6.8	7.5	7.0	6.9

Table 6A

	Math	Span.	Phys.	Chem.	Bio.	His.	Geo.
work	-16.0	-25.8*	4.5	-26.1*	-19.3**	8.2	-5.6
s.e.	12.1	11.4	11.7	10.5	11.6	10.8	10.6
hours	6.0*	8.0*	6.9*	5.4*	5.9*	4.7*	5.2*
s.e.	1.9	1.1	1.1	1.0	1.1	1.1	1.0
eggs	3.9*	2.1	2.8**	-0.5	1.5	2.1	2.3**
s.e.	1.5	1.5	1.5	1.3	1.4	1.4	1.4
f. learn	35.5*	14.9*	19.6*	18.8*	10.9	17.6*	16.8*
s.e.	7.0	6.6	6.8	6.1	6.7	6.3	6.2

Table 6B

	Math	Span.	Phys.	Chem.	Bio.	His.	Geo.
s6	-91.5*	-24.0	-60.7*	-63.3*	-68.9*	-36.7*	-26.0**
s.e.	16.6	15.6	16.1	14.4	15.9	14.8	14.6
s12	-63.3*	-32.4	-45.2*	-3.7	15.9	-23.7	9.9
s.e.	20.0	18.8	19.4	17.4	19.2	17.9	17.6
s16	-45.3*	-25.4	-42.6*	-18.3	-38.4*	-38.8*	-42.1*
s.e.	18.0	17.0	17.5	15.7	17.3	16.1	15.9
s17	-10.0	49.5*	27.8	-12.9	-24.8	-1.7	7.0
s.e.	21.4	20.2	20.8	18.7	20.5	19.1	18.9
s28	65.1*	45.8*	25.4**	-2.4	0.8	53.1*	50.7*
s.e.	15.4	14.5	15.0	13.4	14.8	13.8	13.6
s30	17.6	36.9*	62.7*	1.4	20.9	39.2*	35.2*
s.e.	15.4	14.5	15.0	13.5	14.8	13.8	13.6

Table 6C

Results in table 6A, 6B, 6C; correspond to the OLS estimation.
* significant at 5%
** significant at 10%

Table 7

	Math	Span.	Phys.	Chem.	Bio.	His.	Geo.
cons	872.1*	830.1*	841.2*	805.4*	824.6	828.3*	872.9*
s.e.	42.8	30.2	34.7	36.7	36.7	27.8	31.6
fath. edu	0.7	0.7	2.4**	0.8	0.6	1.0	0.8
s-e	1.5	1.2	1.3	1.3	1.3	1.1	1.2
moth. edu	2.1	4.3*	0.3	2.2**	3.7*	2.8*	1.2
s.e.	1.4	1.2	1.2	1.2	1.3	1.1	1.1
state	0.7	14.8	5.4	7.1	-2.5	-7.9	-29.3*
s.e.	15.1	11.3	12.6	12.9	13.2	10.4	11.7
tele.	-43.7*	6.5	-16.4**	8.3	-1.7	-3.6	3.1
s.e.	10.1	9.5	9.8	8.9	9.7	9.0	9.0
rural	-6.9	-9.9	-11.2	4.3	-6.8	-9.1	1.5
s.e.	11.8	10.2	10.8	10.2	10.9	9.6	10.0
sibl.	-1.0	0.1	-2.1	-3.8*	-0.7	0.2	-0.9
s.e.	1.7	1.6	1.7	1.5	1.6	1.5	1.5
Spanish	17.9**	14.2	19.5**	14.6**	29.6*	28.8*	12.8
s.e.	10.7	9.9	10.2	9.3	10.1	9.4	9.4
Oport.	-51.6	-36.8	-24.7	24.7	-3.4	-17.5	-63.3*
s.e.	43.1	27.5	33.1	36.6	36.1	24.8	30.1

Table 7A

	Math	Span.	Phys.	Chem.	Bio.	His.	Geo.
work	-15.3	-25.6*	4.8	-26.6*	-19.3**	8.5	-4.9
s.e.	12.0	11.3	11.6	10.5	11.5	10.7	10.5
hours	6.0*	8.0*	6.9*	5.4*	5.9*	4.7*	5.2*
s.e.	1.2	1.1	1.1	1.0	1.1	1.1	1.0
eggs	4.0*	2.1	2.8**	-0.5	1.5	2.1	2.4**
s.e.	1.5	1.4	1.5	1.3	1.5	1.4	1.3
f. learn	35.5*	15.0*	19.6*	18.8*	10.9	17.6*	17.0*
s.e.	7.0	6.6	6.7	6.1	6.7	6.2	6.1

Table 7B

	Math	Span.	Phys.	Chem.	Bio.	His.	Geo.
s6	-92.0*	-24.1	-60.9*	-63.0*	-68.9*	-37.2*	-27.0**
s.e.	16.4	15.5	15.9	14.3	15.8	14.7	14.5
s12	-63.6*	-32.5**	-45.4*	-3.2	15.9	-23.8	9.6
s.e.	19.9	18.7	19.2	17.3	19.0	17.7	17.5
s16	-43.7*	-24.8	-41.7*	-19.6	-38.5*	-38.1*	-40.1*
s.e.	17.9	16.7	17.4	15.7	17.2	16.0	15.8
s17	-10.2	49.4*	27.7	-12.5	-24.8	-1.8	6.0
s.e.	21.2	20.0	20.6	18.5	20.4	19.0	18.7
s28	63.2*	45.3*	24.5**	-1.1	0.8	52.4*	48.1*
s.e.	15.4	14.5	14.9	13.4	14.8	13.7	13.5
s30	16.0	36.8*	61.9*	2.5	21.0	38.6*	33.0*
s.e.	15.3	14.5	14.9	13.4	14.7	13.7	13.5

Table 7C

Results in table 7A, 7B, 7C; correspond to the ML estimation with treatment effect.
* significant at 5%
** significant at 10%