Effects on School Enrollment and Performance of a Conditional Cash Transfer Program in Mexico

Pierre Dubois*, Alain de Janvry†, Elisabeth Sadoulet†

This Version: September 2011‡

Abstract

We study the effects of the Mexican conditional cash transfer program Progresa (now renamed Oportunidades) on school enrollment and performance in passing grades. We develop a theoretical framework of the dynamics of the educational process including endogeneity and uncertainty of school performance. It provides predictions for the effect on performance of a cash transfer conditional on school attendance. Using a randomized experiment implemented under Progresa, we identify the effect of the program on enrollment and performance in the first year of the program, before performance-induced dynamic selection took place. We find that the program had a positive impact on school enrollment at all grade levels whereas for performance it had a positive impact at the primary school level but a negative impact at the secondary level. According to our theoretical framework, this can be due to the disincentives created by termination of program benefits after the third year of secondary school.

Key words: education demand, school enrollment, school performance, dynamic decisions, transfer program, Mexico.

JEL Classification: C14, C25, D91, H52, H53, I21, I28, J24.

*Corresponding author:
Toulouse School of Economics (GREMAQ, INRA, IDEI)
Manufacture des Tabacs
21 allée de Brienne
31000 Toulouse France
pierre.dubois@tse-fr.eu

†University of California, Berkeley.

‡We are grateful to Progresa for providing us the data, and Progresa staff of Hidalgo State for guiding us in visiting communities. We thank the Editor Peter Arcidiacono and two anonymous referees as well as Jerome Adda, Orazio Attanasio, Antoine Bommier, Richard Blundell, Luis Braido, Pierre-André Chiappori, Carlos da Costa, Bruno Crépon, Marcelo Fernandes, Denis Fougère, Andrew Foster, Gustavo Gonzaga, Sylvie Lambert, Guy Laroque, Pascal Lavergne, Ethan Ligon, Thierry Magnac, Torsten Persson, Bas Van der Klaauw, Ed. Vytlacil for their comments as well as the seminar participants at CREST Paris, the University of Toulouse, PUC Rio de Janeiro, Getulio Vargas Foundation in Rio de Janeiro, ECARES, Université Libre de Bruxelles, INRA LEA Paris, CIRPEE Université Laval, Québec, Tinbergen Institute, the Institute of International Economic Studies at Stockholm University, University College London, and conferences of Canadian Economic Association Conference in Montreal, Society for Economic Dynamics Conference in Stockholm, Econometric Society European Meeting and European Economic Association Annual Conference in Lausanne, Brazilian Econometric Society Meeting in Porto Seguro.

1 Introduction

In 1998, the Education, Health, and Nutrition Program, known under its Spanish acronym as Progresa and later renamed Oportunidades, was introduced in rural Mexico. The purpose of the program is to create incentives to increase the human capital of children of poor rural households, thus attempting to break the inter-generational inheritance of poverty. To do this, the program provides cash transfers and in-kind benefits to poor households, conditional on the child's school attendance and on regular visits to health centers. On average, these cash transfers represent 22% of the income of beneficiary families. The program has grown rapidly and was covering 2.6 million rural families in extreme poverty in 2000, corresponding to about 40 percent of all rural families in Mexico. At that date, Progresa operated in 50,000 localities in 31 states, with a budget of approximately one billion dollars. In 2008, Oportunidades covered more than 5 million families in both rural and urban areas.

In Mexican rural communities, children tend to join the labor force at early ages, with negative implications on school attendance and performance. High repetition rates further lower the education achievements relative to years of schooling. One of the main objectives of Progresa is to reduce this early labor force participation of children and thereby increase their enrollment and attendance at school. The program includes three closely linked components–education, health, and nutrition– based on the idea that positive interactions between these three components enhance the effectiveness of an integrated program over and above the separate benefits from each component. The educational component constitutes the largest part of the monetary benefits.

The purpose of this paper is to evaluate the impact of Progresa on the educational performance of children. We develop a dynamic model of education demand incorporating the Progresa grants system, and show how it affects not only enrollment decisions but performance at school, a crucial point in the context of high repetition rates. A key feature of this model is the endogenous learning effort. Effort affects performance, which influences school drop out and the final completed education level. Descriptive statistics indeed show that drop out is higher when the child has to repeat a class. The most recent education demand models embody the dynamics and uncertainty associated with wages and returns to schooling as well as liquidity constraints (De Vreyer, Lambert, Magnac, 1999; Magnac and Thesmar, 2002a, 2002b; Cameron and Heckman, 1998, 2001; Cameron and Taber, 2004; Rosenzweig and Wolpin 1995; Eckstein and Wolpin, 1999; Keane and Wolpin, 1997, 2001; Taber, 2001). But most models assume that households can choose with certainty each child's final level of school attainment or at least that the decision to continue updated each year does not involve any uncertainty in grade progression. A few exceptions can be found. For example, Magnac and Thesmar (2002a) use a model where grade completion is stochastic, and show that the rise in educational levels observed in France between 1980 and 1993 was partly due to the decreasing selectivity of the education system. Attanasio, Meghir, and Santiago (2011) study the effect of Progresa on enrolment using a structural model where schooling costs are stochastic and class repetition is allowed. However, they assume that the probability of failing to complete a grade is exogenous and does not depend on effort or on the willingness to continue schooling. Using the same Progresa data, Todd and Wolpin (2006) estimate a structural model of child schooling and fertility, in which similarly the probability of success does not depend on any endogenous effort. Cameron and Heckman (1998) model the transitions from one grade to the next as random processes but without distinguishing whether non progression comes from school drop out or from repetition. To our knowledge, there is no theoretical model where both the endogeneity and uncertainty in successfully passing grades are explicitly modeled. Here, we explicitly take these features into account because both school enrollment and school performance determine educational attainment. In a related study, Behrman, Sengupta, and Todd (2005) estimate school transition matrices by grade and find results quite consistent with ours. Their reduced form specification is however different as they look at transition probabilities from one grade to another conditional on age and do not treat differently students that drop out of school from students that fail. In contrast, we estimate separately a continuation decision and a performance equation. These two approaches are not nested and are equally general.

The theoretical model shows that transfers can have either positive or negative effects on performance. The first order impact of the program on enrollment depends on the size of the current transfer relative to the opportunity cost of time spent at school. The program can also increase the learning effort of children as they want to receive future transfers that increase with grade. On the other hand, program termination may create disincentive effects. The net effect is therefore an empirical issue.

We use data from Progresa's randomized experiment to empirically estimate the effects of the conditional cash transfer on the discrete outcomes of school continuation and performance (success or failure of a grade). The randomized experiment which took place in the middle of a school year, helps solve the usual identification problem of dynamic selection in discrete models with unobserved heterogeneity (Cameron and Heckman, 1998; Ham and Lalonde, 1996). We thus estimate the average program impact at all grade levels in the first year of the program without bias. We find that Progresa had a positive impact on school continuation at all grade levels whereas for performance it had a positive impact at the primary school level but a negative effect at the secondary school level. This is consistent with the findings of Behrman, Sengupta, and Todd (2005) and Schultz (2004).

In section 2, we characterize the related problems of low enrollment and poor school performance in rural Mexico, describe the Progresa transfers with the incentive effects they create, and present the data. In section 3, we develop a life cycle model of education demand that captures how the program design impacts individual education decisions. Section 4 treats of the identification problems. Estimation results are presented in Section 5. Section 6 concludes.

2 Education in Rural Mexico and Progresa

2.1 School Enrollment and Performance

Although educational levels have been improving over time in Mexico, current levels in poor rural communities remain very low, with only 36% of 18 years old having gone beyond primary school. The major breaking point in school attendance occurs at entry in secondary school (Table 1). In primary school, continuation rates reach at least 95% in every grade, with the result that 85% of the children that start primary school complete the cycle. However, only 72.4% of the children that successfully complete primary school enroll in the first year of secondary school. The gender difference is very pronounced at this decisive step, with 75.1% of the boys entering secondary school and only 69.4% of the girls.

Table 1 illustrates the key role that school performance plays on the decision to continue. Throughout primary school continuation rates are higher among those that passed than among those that failed their grade. There is here again a striking discontinuity at entry into secondary school. The performance rate is the lowest in the first year of secondary school and dropout rates after a first year of trying secondary school without success are very high. In the last year of each cycle we observe very large re-enrollment rates upon failure, suggesting important incentives to complete a cycle.

Table 1: Continuation and Performance Statistics

(% in 1998 for children at school in 1997)	Overall Continuation		Performance (grade success)		Continuation among those that passed		failed	
Grade attended in 1997	Boys	Girls	Boys	Girls	Boys	Girls	Boys	Girls
Primary school								
2	98.5	97.5	76.2	77.9	98.5	97.1	98.5	98.8
3	96.9	96.8	77.9	78.1	96.9	97.8	96.8	92.9
4	96.7	96.0	78.5	80.4	97.2	96.9	95.2	92.2
5	94.9	95.9	83.7	85.1	95.6	96.8	91.0	91.0
6	75.1	69.4	84.1	84.7	73.5	65.9	83.9	89.3
Secondary school								
1	82.3	77.5	74.2	67.8	94.3	94.9	47.5	41.1
2	95.5	95.1	84.4	82.1	96.2	96.1	91.8	90.2
3	57.3	59.4	80.7	75.3	36.9	49.0	91.0	91.8

Note: These descriptive statistics are for the whole population ("poor" and "non poor") of children in control villages (i.e., villages where Progresa is not implemented).

In conclusion, there is a clear school continuation problem, especially at entry into secondary school, and it is intertwined with a problem of performance in passing grades.

2.2 Progresa's Incentive Scheme

Progresa is targeted at poor families and has three components: health, nutrition, and education. The health component offers basic health care to all family members. The nutrition component includes a fixed monetary transfer for improved food consumption, as well as nutritional supplements targeted at all children under the age of two, malnourished children under the age of five, and pregnant and breast-feeding women. Families must complete a schedule of visits to health care facilities in order to receive monetary support for improved nutrition. Education is, however, by far the program's most important component in terms of cash transfers. It consists in payments to families with children attending school between the third grade of primary and the third grade of secondary. The conditionality requires presence at school in at least 85% of school days, i.e., no more than 3 absences a month.[1] After three years in the program, families may renew their status as beneficiaries, subject to reevaluation of their socio-economic condition. The level of the transfers (see Table 2) increases as children progress to higher grades in order to match the rising income they would contribute to their families if they were working (Progresa, 2000). The transfers are slightly higher for girls than for boys in secondary school.

In addition, program rules impose an upper limit to the total cash transfer a household can receive. Details about this rule are explained and used in section 4.3.

[1] Another rule implies that students lose eligibility if they repeat a grade twice. However, the data do not give information on the student's past schooling history before the October 1997 baseline survey. Moreover, the rule of grade repetition does not appear to have been enforced and surely not the first year of the program to students already repeating. We thus ignore such rule for a simpler modeling and in order to be consistent with the actual implementation of Progresa.

Table 2: Monthly Progresa Transfers in Pesos

Educational Grant by Student	1997		1998		1999	
Primary School	Boys	Girls	Boys	Girls	Boys	Girls
1^{st} and 2^{nd} year	0	0	0	0	0	0
3^{rd} year	60	60	70	70	80	80
4^{th} year	70	70	80	80	95	95
5^{th} year	90	90	100	100	125	125
6^{th} year	120	120	135	135	165	165
Secondary School	Boys	Girls	Boys	Girls	Boys	Girls
1^{st} year	175	185	200	210	240	250
2^{nd} year	185	205	210	235	250	280
3^{rd} year	195	225	220	255	265	305
Further Schooling	0	0	0	0	0	0
Cash Transfer for Food	90		100		125	
Household level maximum benefit	550		625		750	

Note: Nominal values corresponding to the second semester of the year (changes occur every semester).

Approximate exchange rate: 10 Pesos = 1 US$.

Given these program rules, several incentive mechanisms can affect the behavior of beneficiary households. All transfers, conditional or unconditional, create an income effect.[2] Higher school attendance may increase students' knowledge and reduce grade repetition.[3] The conditionality on school attendance creates both static and dynamic incentives to enroll. The static incentive relates to the current transfer payment, which reduces the foregone income in going to school. The option of receiving future transfers if one stays in school creates the dynamic effect. Rising transfers with grade level further enhance the dynamic incentives. They create an incentive to pass grades and thus to perform better at school, in addition to the threat of losing eligibility after two repetitions. A possible negative effect of program termination may however appear since, in the third year of secondary school, students may prefer to repeat the grade rather than lose the transfer. This disincentive effect could also increase repetition before the last grade of eligibility in order to increase the number of years a student stays in the program.[4]

[2] To the extent that preference for and performance in school increase with income, transfers will have a positive effect. This is all the more true if the household faces short term liquidity constraints and school attendance entails monetary costs such as transportation.

[3] At the individual level, higher school attendance is expected to improve learning. However, there may be negative externalities on those students which, in any case, would have attended school regularly if the increased number of children in a classroom lowers school quality. As the data do not indicate the exact class attendance, we cannot test for the presence of these externalities.

[4] To get an order of magnitude, in rural Mexico, the average daily wage of a 16-18 years old boy with completed junior high school in the sample was 25 pesos in 1997. A full time work of 20 days per month would generate an income of 500 pesos per month, compared to a maximum of 255 pesos from Progresa transfers. Therefore, heterogeneity of individuals and of labor market uncertainty implies that some students may prefer to repeat the grade rather than look for a job.

2.3 Randomized Experiment and Average Effects

Progresa started operating in poor rural communities (defined on the basis of a national marginality index using the 1995 population census) with sufficient access to primary school and primary care facilities. The program was gradually rolled out and partial coverage in the first years was used to facilitate evaluation with a randomized experiment. A subset of 506 of the 50,000 eligible communities was selected to participate in the evaluation. 320 of these communities were randomly assigned to constitute the treatment group where Progresa was implemented starting in May 1998, while the remaining 186 communities formed the control group where Progresa would be introduced three years later (Behrman and Todd, 1999). These experimental communities are located in seven states: Guerrero, Hidalgo, Michoacan, Puebla, Queretaro, San Luis Potosi, and Veracruz. On average, 78% of the households in the selected communities were deemed in poverty and hence eligible for the program (the poverty status of households was established prior to the start of the program using a household census conducted in October 1997, see Skoufias, Davis, and Behrman, 1999). All households (eligible and not) of both types of communities were then surveyed twice a year during the three years of the evaluation. For our analysis, we use data from the baseline survey in October 1997 and the follow up surveys in May and October 1998. We thus have information on enrollment during school years 1997-98 and 1998-99, and on performance in school during 1997-98 on some 13,900 children from poor households. Because transfers are generous, almost all eligible families chose to participate (97%). The average household size is around 7. 15% of household heads have an educational level less than primary school, 30% completed primary school, and 52% completed secondary school. We first look at instructive descriptive statistics contrasting treatment and control communities.

Table 3: Continuation and Performance

Grade attended in 1997	Continuation Control	Continuation Treatment	Performance Control	Performance Treatment
Primary school				
2	0.982	0.991*	0.764	0.778
3	0.967	0.987*	0.775	0.828*
4	0.964	0.984*	0.788	0.833*
5	0.963	0.978*	0.846	0.864
6	0.703	0.824*	0.837	0.858
Secondary school				
1	0.803	0.898*	0.728	0.760
2	0.957	0.972	0.811	0.860*
3	0.543	0.591	0.755	0.721

Note: Overall rates for students of poor families in treated and control communities. Performance is measured as the difference between the grade completed in October 1997 and the grade completed in

October 1998. It is 1 if passed and 0 if failed. School continuation is defined as being enrolled at school in 1998 while already at school in October 1997. It is 1 if the student continues schooling and 0 if drops out. Statistical significance of equality of means has been tested and rejection at the 5% level if denoted with * in the treatment column.

Table 3 provides statistics that motivate the education demand model developed in the next section and the empirical econometric estimates. School continuation rates are consistent with those reported by Schultz (2000, 2004). Comparison between control and treated communities show that Progresa benefits seem to have increased school continuation at all grade levels and particularly at the end of primary school and first year of secondary school. Results on performance are showing an increase at primary school and secondary school levels except for the third year of secondary school which is the last year of program eligibility where repetition seems to have increased in Progresa villages, something consistent (but not identical) with Behrman, Sengupta, and Todd (2005).

3 A Dynamic Educational Model with Schooling, Effort, and Performance

We construct a dynamic schooling model showing how a conditional cash transfer affects both enrollment and learning. We assume that the household decision maker's objective is to maximize the child's net expected income. In this calculus, education has a cost which includes the opportunity cost of the time spent studying instead of working. The return to schooling and the opportunity cost of attending school increase with the individual's completed school grade, and successfully completing grades is uncertain and endogenous (it depends on a learning effort variable). We explicitly include in the model conditional cash transfers. The program may impact learning effort, performance, or grade progression through other channels than the cash transfers conditional on school attendance (e.g., better health, better nutrition, better attendance, or classroom composition and peer effects) but, as these transfers are Progresa's main and most important component, we choose to model them explicitly abstracting from other aspects of the program that will have to be kept in mind when interpreting the results.

A child of gender g who has completed grade l at the beginning of an academic year is assumed to be automatically accepted in grade $l+1$ if he enrolls at school. If the household is eligible, he is then entitled to an educational transfer denoted $\tau(l, g, p)$. In our sample, p is equal to 1 in randomly selected treatment villages and 0 otherwise (with $\tau(l, g, 0) \equiv 0$). Note that with the cap on total household transfer, direct incentives to attend school are a function of family structure, giving us a possible variation in the value of transfers across children (as will be used in section

4.3). Let s be a variable equal to one if the child attends school and zero otherwise. Let π, the educational performance of the child, be a function of his school level l, and an individual learning effort e: $\pi(l, e)$. This effort variable is meant to represent individual actions of the student such as attention in class, being at school on time, and studying at home. As educational learning and skills are not perfectly observable by the teacher, we assume that the student will complete grade $l+1$ if and only if $s = 1$ and $\pi(l, e) \geq \varepsilon$, where ε is a random variable with c.d.f. F and p.d.f. f. π depends on l because the level of effort required to pass varies with grade level. This function depends on the selectivity of the educational system as set by government (that is how hard exams are to pass). The function π can also depend on characteristics x of the student, but we don't need to explicitly introduce them in the theoretical model as long as they are exogenous.

Grade progression from year t to year $t+1$ is then determined by the following rule:

$$\begin{aligned} l_{t+1} &= l_t + 1 \qquad \text{if } s_t = 1 \quad \text{and} \quad \pi(l_t, e_t) \geq \varepsilon_t \\ &= l_t \qquad \text{if } s_t = 0 \quad \text{or} \quad \pi(l_t, e_t) < \varepsilon_t \end{aligned} \tag{1}$$

with the following assumptions:

Assumption 1 The probability of success $P(l_{t+1} = l_t + 1 | e_t, s_t = 1) = F \circ \pi(l_t, e_t)$ is increasing and concave in effort e_t.

This assumption is satisfied when the performance function $\pi(l, e)$ is increasing and concave in e and the random terms ε are i.i.d. across individuals and periods and their c.d.f. F is concave.

We assume that a person with gender g and completed grade l, is able to work (either on farm, at home, or outside) and gets earnings $w(g, l)$ (again the model could be written with $w(x, g, l)$ where individual characteristics x affect earnings).

Assumption 2 The earnings function $w(g, l)$ is increasing in the acquired level of education l.

We assume that the cost for a child of going to school, denoted $c(e)$, depends on the learning effort e (plus the cost of transportation and other costs associated with enrollment).

Assumption 3 The cost function $c(e)$ is increasing and convex in e, the level of learning effort at school.

Then, sending a child to school in year t costs $c(e_t) - \tau(l_t, g, p)$, while not sending him generates earnings $w(g, l_t)$, the opportunity cost of school enrollment. Assuming that the household's decision process results in maximization of the child's intertemporal expected benefits $V(l_t, g, p, s_t)$, the value of enrolling a child at the beginning of year t ($s_t = 1$) or of not enrolling him ($s_t = 0$) knowing his completed grade l_t, gender g, and eligibility p can be written recursively as follows:

$$\begin{aligned} V(l_t, g, p, 1) &= \max_{e_t}\{\tau(l_t, g, p) - c(e_t) + \beta E[\max_{s_{t+1}\in\{0,1\}} V(l_{t+1}, g, p, s_{t+1}) \mid s_t = 1]\} \quad (2) \\ V(l_t, g, p, 0) &= w(g, l_t) + \beta E[\max_{s_{t+1}\in\{0,1\}} V(l_{t+1}, g, p, s_{t+1}) \mid s_t = 0] \quad (3) \end{aligned}$$

with β the discount factor and l_{t+1} following the rule in (1).

Because of the uncertainty of grade progression, parents revise their expected optimal choice at the beginning of each school year.

The value function for a child of education l_t, gender g, and eligibility p can be written:

$$\phi(l_t, g, p) = \max_{s_t\in\{0,1\}} V(l_t, g, p, s_t)$$

Substituting in expressions (2) and (3) gives

$$\begin{aligned} V(l_t, g, p, 1) &= \max_{e_t}\{\tau(l_t, g, p) - c(e_t) + \beta E[\phi(l_{t+1}, g, p) \mid s_t = 1]\} \\ V(l_t, g, p, 0) &= w(g, l_t) + \beta E[\phi(l_{t+1}, g, p) \mid s_t = 0] = w(g, l_t) + \beta\phi(l_t, g, p) \end{aligned}$$

because $P(l_{t+1} = l_t + 1 \mid s_t = 0) = 0$. This implies

$$\phi(l_t, g, p) = \max\{\tau(l_t, g, p) + \max_{e_t}\{\beta E[\phi(l_{t+1}, g, p) \mid s_t = 1] - c(e_t)\}, w(g, l_t) + \beta\phi(l_t, g, p)\} \quad (4)$$

We can show the following proposition:

Proposition 1 *The function ϕ defined by the Bellman equation (4) exists and is unique.*

Proof. See Appendix C.1. ■

Intuitively, we expect the value function ϕ to be increasing with completed grade. In Appendix B, we give sufficient conditions on the primitives of the model that ensure that it is the case: If individuals are sufficiently patient ($\beta \geq 1/2$) and transfers are increasing with grade (or not decreasing too much, i.e., transfers are not too large when the program ends) and not too large compared to the potential wage, then the endogenous value function ϕ is increasing with completed grade. In contrast, ϕ will be clearly decreasing with completed grade if the lifetime wage did not respond to education, and the child was losing a very high transfer in enrolling beyond the upper grade covered by the program. However, rather than assuming these "reasonable" sufficient conditions (likely to be true in the case of Progresa), we will simply impose the assumption that $\phi(l, g, p)$ is increasing in l. Thus, in the rest of this paper, we will assume that:

Assumption 4 The value function $\phi(l, g, p)$ is always increasing with completed grade l.

3.1 Program Impact on Effort and Performance

We could consider, as often done, that learning effort depends on children's exogenous characteristics and cannot be adjusted once presence at school is required. Then, effort would have to be considered exogenous in our model. It is however clear that higher returns to education (in a very broad sense) provide an incentive for students to study and learn more. Introducing an endogenous learning effort, maximization of the value function implies that learning effort is chosen conditional on enrollment so as to maximize $\tau(l_t, g, p) - c(e) + \beta E[\phi(l_{t+1}, g, p) \mid s = 1]$. Assuming that there is no fixed cost to effort and that the cost of zero effort is zero ($c(0) = 0$), then we obtain the following proposition[5]:

Proposition 2 *The learning effort e^* is strictly positive if and only if $\phi(l+1, g, p) > \phi(l, g, p)$ and satisfies the first order condition*

$$\beta[\phi(l+1, g, p) - \phi(l, g, p)] f \circ \pi(l, e^*) \frac{\partial \pi}{\partial e}(l, e^*) = c'(e^*) \tag{5}$$

Proof. See Appendix C.2. ■

Since the performance technology depends on grade (difficulty of tests, system selectivity, etc.), proposition 2 implies that effort at school can be highly non-linear and non-monotonic in the grade level. It can be either increasing or decreasing with grade and thus the expected performance at school can also be either increasing or decreasing with grade.

These results call for taking into account heterogeneity of treatment effects, as shown by the following proposition:

Proposition 3 *Treatment raises effort and performance at school in terms of probability to succeed in a given grade l_t if $\phi(l_t + 1, g, 1) - \phi(l_t, g, 1) > \phi(l_t + 1, g, 0) - \phi(l_t, g, 0)$ and reduces effort and expected performance if $\phi(l_t + 1, g, 1) - \phi(l_t, g, 1) < \phi(l_t + 1, g, 0) - \phi(l_t, g, 0)$.*

Proof. As shown in Appendix C.2, if $\phi(l_t + 1, g, p) - \phi(l_t, g, p) \leq 0$, then $e_t^* = 0$. If $\phi(l_t + 1, g, p) - \phi(l_t, g, p) > 0$, then e_t^* satisfies (5). For notational ease we use Δ for the discrete difference in p, that is for a function $H(p)$, $\Delta H(p) = H(1) - H(0)$. With assumptions 1 and 3, the implicit function theorem implies that e_t^* is increasing in p if $\Delta[\phi(l_t + 1, g, p) - \phi(l_t, g, p)] \geq 0$ and decreasing in p otherwise. As $\pi(l, e)$ is an increasing function of e, the same applies for performance. ■

This proposition indicates that the effect of the program on learning and performance at school will depend on whether the value of getting an additional year of education is higher for treated or untreated students. Its sign is ambiguous and depends on the effect of the program on the

[5] With a non zero "fixed cost" of effort, the optimal effort is either zero or is strictly positive and satisfies (5).

curvature of $\phi(l, g, p)$ with respect to l. Note that the program can have a positive impact on effort at a grade l while it has a negative impact at another lower or higher grade l'.

To get an intuition on the expected sign of this expression, consider the case where the optimal path for the child with completed grade l includes completed grade $l+1$ the following year. The value function at completed grade l is thus equal to: $\phi(l_t, g, p) = \tau(l_t, g, p) - c(e) + \beta\phi(l_t + 1, g, p)$ and then $\phi(l_t + 1, g, p) - \phi(l_t, g, p) = -\tau(l_t, g, p) + c(e) + (1 - \beta)\phi(l_t + 1, g, p)$. Taking the difference in p we have

$$\Delta[\phi(l_t + 1, g, p) - \phi(l_t, g, p)] = -\tau(l_t, g, p) + (1 - \beta)\Delta[\phi(l_t + 1, g, p)].$$

The conditional cash transfer program reduces the net cost of the current school year by τ and raises the next year value function $\phi(l_t + 1, g, p)$ by the direct and indirect value of potential transfers in upper grades. Indirectly therefore, all the parameters of the model enter this expression, such as cost of effort, earnings function, etc., corresponding to grades above $l_t + 1$. One dimension of particular interest however is due to the limited length of the program. For children nearing the end of the program, $\Delta[\phi(l_t + 1, g, p)]$ is small, in fact null for a child with completed grade $l_t = 2^{nd}$ year of secondary school, since there will be no transfer beyond the current year. Hence, $\Delta[\phi(l_t + 1, g, p) - \phi(l_t, g, p)]$ will be either small or negative. Progresa is thus expected to have either a small or even a negative effect on effort. In contrast, at lower grade levels, the program raises $\phi(l_t + 1, g, p)$ by the value of many potential years of transfer, while reducing the cost of the current year in schooling by the same τ. Progresa is thus expected to induce an increase in effort unless the child is so patient that he would rather enjoy adding one more year of transfer now and pushing down his future plan by one year. We will test this heterogeneity of impact across grades in the empirical section.

Note also that the previous proposition gives us an empirical test of whether the learning effort is actually endogenous which is of great importance for education policies. Under the assumption that treatment does not directly affect the performance and evaluation technologies π and F, if treatment affects the probability to pass a given grade, it means that the learning effort is endogenous.

3.2 Program Impact on the Enrollment Decision

The enrollment decision is taken by comparing the values of going and not going to school. Define the decision to enroll the child by $s_t = 1_{\{v(l_t,g,p)\geq 0\}}$ where $v(l_t, g, p) = V(l_t, g, p, 1) - V(l_t, g, p, 0)$ is the difference between the two conditional value functions.

The following proposition shows the derivatives of $v(l_t, g, p)$ with respect to the program treatment p which represents the effect of treatment on the propensity to choose schooling over working.

Proposition 4 *The program impact on the value of going to school compared to not going is:*

$$\begin{aligned}\Delta v(l_t, g, p) &= \tau(l_t, g, 1) + P(l_{t+1} = l_t + 1 \mid s_t = 1)\beta\Delta\left[\phi\left(l_t + 1, g, p\right) - \phi\left(l_t, g, p\right)\right] \\ -\Delta h(P(l_{t+1} &= l_t + 1 \mid s_t = 1)) + \beta\left[\phi\left(l_t + 1, g, p\right) - \phi\left(l_t, g, p\right)\right]\Delta P(l_{t+1} = l_t + 1 \mid s_t = 1)\end{aligned}$$

where $h(.)$ is a positive and increasing function equal to the inverse in effort of the probability of success given effort.

Proof. See Appendix C.3. ■

Proposition 4 shows that $\Delta v(l_t, g, p)$ depends on the direct incentives to go to school provided by the educational transfer, $\tau(l_t, g, 1) \geq 0$, and on the discounted expected marginal value of program eligibility composed of several terms:

- The probability of successfully completing the current grade times the increase in getting one additional year of education provided by the treatment p (which implicitly includes the future years benefits of the program): $P(l_{t+1} = l_t + 1 \mid s_t = 1)\beta\Delta[\phi(l_t + 1, g, p) - \phi(l_t, g, p)]$.
- The marginal value of getting an additional year of education times the change in the probability of succeeding: $\beta\left[\phi\left(l_t + 1, g, p\right) - \phi\left(l_t, g, p\right)\right]\Delta P(l_{t+1} = l_t + 1 \mid s_t = 1)$.
- The change in the cost of effort or equivalently the change in an increasing function of the probability to pass $\Delta h(P(l_{t+1} = l_t + 1 \mid s_t = 1))$.

Moreover, according to proposition 3, $\Delta P\left(l_{t+1} = l_t + 1 \mid s_t = 1\right)$ and $\Delta[\phi(l_t+1, g, p) - \phi(l_t, g, p)]$ are of the same sign. This model clearly shows the implications of the program on the value for children of going to school compared to not going. In particular, it helps explain that the incentives provided by the program depend not only on reduction of the opportunity cost of schooling by the conditional transfers but also on the additional value provided by the expectation of receiving transfers the year after and the expected value of being more educated. Therefore, the program impact has no reason to be simply proportional to the transfers received. According to this model, the program's treatment effect should be heterogeneous not only across individuals with different grades and genders but also for example across all characteristics that affect future wages.

Another implication is that the incentive to go to school represented by $\Delta v(l_t, g, p)$ depends on the probability of grade progression, on the cash transfer corresponding to the current grade, and on transfers for higher but not for lower grades.

Note also that even if the transfer function for some grade l and gender g is zero, $\tau(l, g, 1) = \tau(l, g, 0) = 0$, we still have $\Delta v(l, g, p) \neq 0$ if for some grade $l' > l$, $\tau((l', g, 1) > 0$. Because of the expected benefit from transfers in higher grades, the cash transfer generates incentives in favor of schooling even if the student is not entitled to receive any grant in his current grade. In the particular case of Progresa, this suggests the existence of an incentive to schooling in the first and second years of primary school even though students receive nothing in their current grade.

4 Identification and Econometric Evaluation

The theoretical model developed in the previous section has testable implications regarding the impact of a conditional cash transfer on enrollment decisions and performance outcomes. We now turn to the estimation of passage probabilities from grade to grade $P(l_{t+1} = l_t + 1 \mid s_t = 1)$ and continuation decision $P(s_{t+1} = 1 \mid s_t = 1)$ for the year 1997-98.[6]

4.1 Econometric Specification

To simplify notations, the theoretical model did not explicit exogenous characteristics that could affect the wage or costs of schooling. Adding now these characteristics x_t to the grade progression model (1), we have: $P(l_{t+1} = l_t + 1 \mid s_t = 1) = F \circ \pi(x_t, l_t, e_t^*)$ where e_t^* is endogenously determined and depends on x_t, l_t, g, and p. Therefore, we specify the grade progression as:

$$P(l_{t+1} = l_t + 1 \mid s_t = 1) = \varphi(\theta_1 p + \theta_2 \tau(l_t, g, p).p + X_t \gamma_{l_t,g}) \tag{6}$$

where $\gamma_{l,g}$, θ are vectors of parameters specific to grade l and gender g, X_t is a vector of exogenous variables (including x_t, l_t, g), and φ is a c.d.f. (for example logistic or normal).[7][8]

The reduced form of the model does not allow decomposing the effect of the program into each incentive component identified in the theoretical model. However, it allows to evaluate the total program impact on enrollment and performance at school and is more robust to misspecification than a structural model would be.

We know from the model that the decision to continue schooling is given by $s_t = 1_{\{v(l_t,g,p) \geq 0\}}$ and thus that the propensity to continue school is affected by the quantity $\Delta v(l_t, g, p)$. As shown by Proposition 4, $\Delta v(l_t, g, p)$ is a function among other things of the probability to pass $P(l_{t+1} = l_t + 1 \mid s_t = 1)$. Consequently, when specifying the reduced form equation for $P(s_{t+1} = 1 \mid s_t = 1)$, we cannot impose exclusion restrictions where some exogenous determinants of the probability to pass would not affect the continuation probability, while the contrary could be possible. In practice,

[6] A previous longer version (Dubois, de Janvry, and Sadoulet 2007) of this paper presents a semi-structural approach with parametric identification conditions that provide consistent results with those in this paper and additional insights not mentioned here.

[7] All estimations will be based on parametric maximum likelihood. The semi-parametric identification of binary choice models is possible with some assumptions like location and scale normalization (Manski, 1985 and 1988) but given the number of explanatory variables used in our regressions we will use simple parametric estimation methods.

[8] This specification is consistent with the theoretical model for example (but not only) if the c.d.f. F of ε is normal or logistic and the performance function π is a linear index of its arguments.

we specify the school continuation probability in a reduced form as follows

$$P(s_{t+1} = 1 \mid s_t = 1) = \varphi(\alpha_1.p + \alpha_2 \tau\left(l_{t+1}, g, p\right).p + X_t \delta_{l_t,g}) \tag{7}$$

Since the design of the program is such that transfers are gender and grade specific, coefficients α_1 and α_2 will be allowed to be gender and grade specific. Thus, α_2 will be identified only if transfers $\tau\left(l_{t+1}, g, p\right)$ vary for a given grade and gender in treated villages.

4.2 Identification of the Program Impact and the Dynamic Selection Problem

As shown by Cameron and Heckman (1998), the estimation of school transition models faces a problem of dynamic selection bias. Even if unobserved factors entering the school transition model are distributed independently of observable characteristics in the population enrolling in the first year of primary school (for example), the distribution of unobserved characteristics of students in the second year of primary school will be truncated and not independent of the distribution of observable characteristics. Here, this difficulty certainly affects estimation of the probabilities to enroll at school and to successfully pass a grade. However, with randomization of treatment, and a program that started in the middle of a school year, evaluation of the average program impact will not be biased by this dynamic selection problem in the first year of the program. We explicitly formulate the necessary assumptions for identification, and establish the relationship between randomization and the dynamic selection problem.

Grade transition probabilities conditional on the vector of observables $\omega_{t+1} = (X_t, l_{t+1}, s_t)$, the treatment dummy $p \in \{0, 1\}$, and unobserved characteristics $\widetilde{\theta}$ can be written[9]

$$E(s_{t+1} \mid \omega_{t+1}, p, \widetilde{\theta}) = \psi(\omega_{t+1}, p, \widetilde{\theta})$$

where $\psi\left(.\right)$ is a real valued function. It is to be noted that, with these notations, we have heterogenous treatment effects $\Delta\psi(\omega_{t+1}, p, \widetilde{\theta})$.

As $\widetilde{\theta}$ is unobserved, we cannot identify $E(s_{t+1} \mid \omega_{t+1}, p, \widetilde{\theta})$ but we are interested in the average $E(s_{t+1} \mid \omega_{t+1}, p) = E_{\widetilde{\theta}}[E(s_{t+1} \mid \omega_{t+1}, p, \widetilde{\theta})]$. The parameters of interest that we would like to identify are the average program impact

$$E_{\widetilde{\theta}}[\Delta E(s_{t+1} \mid \omega_{t+1}, p, \widetilde{\theta})]$$

and the average of the effects of some covariates ω_{t+1} on treatment effects

$$E_{\widetilde{\theta}}[\frac{\partial}{\partial \omega_{t+1}} \Delta E(s_{t+1} \mid \omega_{t+1}, p, \widetilde{\theta})]$$

[9] Note that $E(s_{t+1} \mid \omega_{t+1}, p, \widetilde{\theta}) = P(s_{t+1} = 1 \mid \omega_{t+1}, p, \widetilde{\theta})$.

Cameron and Heckman (1998) showed that even if the distribution of $\widetilde{\theta}$ is independent of ω_0 ($\widetilde{\theta} \perp\!\!\!\perp \omega_0$), this random effect assumption for the initial schooling stage will not be true for the subsequent ones because of the selection of students; that is, in general $\widetilde{\theta} \not\perp\!\!\!\perp \omega_{t+1}$. This dynamic selection bias implies that

$$\frac{\partial}{\partial \omega_{t+1}} E\left(s_{t+1} \mid \omega_{t+1}, p\right) = \frac{\partial}{\partial \omega_{t+1}} [E_{\widetilde{\theta}} \psi(\omega_{t+1}, p, \widetilde{\theta})] \neq E_{\widetilde{\theta}}[\frac{\partial}{\partial \omega_{t+1}} \psi(\omega_{t+1}, p, \widetilde{\theta})]$$

The value $\frac{\partial}{\partial \omega_{t+1}} E\left(s_{t+1} \mid \omega_{t+1}, p\right)$ is thus a biased estimator of $E_{\widetilde{\theta}}[\frac{\partial}{\partial \omega_{t+1}} \psi(\omega_{t+1}, p, \widetilde{\theta})]$ because the derivative of the average $E_{\widetilde{\theta}} \psi(\omega_{t+1}, p, \widetilde{\theta})$ is not equal to the average derivative. The bias being difficult to sign and quantify *a priori* (as in Cameron and Heckman, 1998), a solution is then to model the unobserved component $\widetilde{\theta}$, for example by using the Heckman and Singer (1984) technique of introducing an arbitrary discrete non-parametric distribution for $\widetilde{\theta}$.

However, as proposition 5 below shows, we do not encounter the same problem when evaluating the average program impact

$$\Delta E(s_{t+1} \mid \omega_{t+1}, p) = E(s_{t+1} \mid \omega_{t+1}, p=1) - E(s_{t+1} \mid \omega_{t+1}, p=0)$$

Actually, randomization implies that treatment is orthogonal to observed and unobserved characteristics

$$p \perp\!\!\!\perp (\widetilde{\theta}, \omega_{t+1}) \tag{8}$$

which implies (9) that can be used in the following proposition.

Proposition 5 *If treatment $p \in \{0, 1\}$ is orthogonal to the distribution of unobserved characteristics conditional on observables ω_{t+1} that is*

$$p \perp\!\!\!\perp \widetilde{\theta} \mid \omega_{t+1} \tag{9}$$

then

$$\Delta E\left(s_{t+1} \mid \omega_{t+1}, p\right) = \Delta[E_{\widetilde{\theta}} \psi(\omega_{t+1}, p, \widetilde{\theta})] = E_{\widetilde{\theta}}[\Delta \psi(\omega_{t+1}, p, \widetilde{\theta})] \tag{10}$$

Proof. Proof in Appendix C.4. ■

This proposition shows that the parameters of interest $E_{\widetilde{\theta}|\omega_{t+1}}[\Delta \psi(\omega_{t+1}, p, \widetilde{\theta})]$ and $E_{\widetilde{\theta}|\omega'_{t+1}}[\Delta \psi(\omega'_{t+1}, p, \widetilde{\theta})]$ are identified for any ω_{t+1} and ω'_{t+1} and thus their difference.

However, we could consider that this is not the parameter of interest and rather that we would like to identify the structural parameter $E_{\widetilde{\theta}|\omega_{t+1}}[\Delta \psi(\omega_{t+1}, p, \widetilde{\theta}) - \Delta \psi(\omega'_{t+1}, p, \widetilde{\theta})]$ which tells us how average treatment effects $\Delta \psi(\omega_{t+1}, p, \widetilde{\theta})$ vary with ω_{t+1} when we average over some constant distribution of unobservables $\widetilde{\theta}$ given ω_{t+1}.

We have the following Proposition:

Proposition 6 *For any ω_{t+1}^k, the average treatment effect $E_{\widetilde{\theta}}[\Delta \frac{\partial}{\partial \omega_{t+1}^k} \psi(\omega_{t+1}, p, \widetilde{\theta})]$ is identified and equal to $\Delta \frac{\partial}{\partial \omega_{t+1}^k} E\left(s_{t+1} \mid \omega_{t+1}, p\right)$ if one of the following condition is satisfied*[10]

$$\frac{\partial}{\partial \omega_{t+1}^k} d\lambda(\widetilde{\theta} \mid \omega_{t+1}) = 0 \tag{11}$$

or

$$\int E(s_{t+1} \mid \omega_{t+1}, 1, \widetilde{\theta}) \frac{\partial}{\partial \omega_{t+1}^k} d\lambda(\widetilde{\theta} \mid \omega_{t+1}) = \int E(s_{t+1} \mid \omega_{t+1}, 0, \widetilde{\theta}) \frac{\partial}{\partial \omega_{t+1}^k} d\lambda(\widetilde{\theta} \mid \omega_{t+1}) \tag{12}$$

Proof. See Appendix C.5. ■

Condition (11) means that the distribution of $\widetilde{\theta}$ does not depend on ω_{t+1}^k i.e. that there is no dynamic selection bias in the direction of ω_{t+1}^k. Condition (12) means that the marginal treatment effect[11] $\Delta E(s_{t+1} \mid \omega_{t+1}, p, \widetilde{\theta})$ averages to zero when integrating with respect to $\frac{\partial}{\partial \omega_{t+1}^k} d\lambda(\widetilde{\theta} \mid \omega_{t+1})$ which is always the case if $\Delta E(s_{t+1} \mid \omega_{t+1}, p, \widetilde{\theta})$ is constant across $\widetilde{\theta}$ because $\int \frac{\partial}{\partial \omega_{t+1}^k} d\lambda(\widetilde{\theta} \mid \omega_{t+1}) = 0$ since $\int d\lambda(\widetilde{\theta} \mid \omega_{t+1}) \equiv 1$. Therefore, this is always true if the average treatment effect $\Delta E(s_{t+1} \mid \omega_{t+1}, p, \widetilde{\theta})$ does not depend on ω_{t+1}.**

If these conditions are not satisfied, it means that the identified parameter $E\Delta\psi(\omega_{t+1}, p) - E\Delta\psi(\omega'_{t+1}, p)$ will tell us how average treatment effects $\Delta\psi(\omega_{t+1}, p, \widetilde{\theta})$ vary with ω_{t+1} but taking average over the distribution of unobservables $\widetilde{\theta}$ which also varies with ω_{t+1}.

In the present case, neither assumption (11) nor (12) have to be valid given the randomization procedure. (11) will be wrong as soon as there is some dynamic selection which is likely in education transition models and (12) is unlikely to happen as soon as the treatment effect $\Delta E(s_{t+1} \mid \omega_{t+1}, p, \widetilde{\theta})$ depends on covariates ω_{t+1}. Therefore, the randomization process only insures identification of the average impact $\Delta E(s_{t+1} \mid \omega_{t+1}, p)$. The same argument can be applied to the performance probability $P\left(l_{t+1} = l_t + 1 \mid s_t = 1\right)$. The randomization condition (9) is sufficient to ensure that the dynamic selection bias present in the estimation of the conditional probabilities $P\left(s_{t+1} = 1 \mid s_t = 1\right)$ and $P\left(l_{t+1} = l_t + 1 \mid s_t = 1\right)$ will be the same for treated and untreated sample and will then cancel out in the estimation of the program impact.

Condition (8) that the joint distribution of observables and unobservables is independent of treatment (i.e., that it is the same across treated and control samples) is not testable, but an implication of it on the marginal distribution of observables can be checked (and empirically validated for data in 1997 by Behrman and Todd (1999) and Schultz (2004)).While we can safely assume that randomization of program placement in the case of Progresa is such that condition (8) is true at the beginning of the program in 1997, we can expect that this will not be the case afterwards. Since the

[10] The notation λ is used to designate cumulative distribution functions. For example $\lambda(\widetilde{\theta} \mid \omega_{t+1})$ is the c.d.f. of $\widetilde{\theta}$ conditional on ω_{t+1}.

[11] The analogy with marginal treatment effects of Heckman and Vytlacil (1999,2001) is due to the fact that the unobservable may affect the probability to be treated through the dynamic selection.

program impact is non zero (as empirical results will confirm), the dynamic selection bias will not cancel out across treatment and control groups. Therefore, probability estimates of the program impact for example between 1998 and 1999 will be biased by a dynamic selection bias due to the impact of the program. For example, if the program has a positive effect on the propensity to continue school, it can select individuals with (on average) lower unobserved factors also causing an increase in the individuals' propensity to go to school (like unobserved ability). This in turn would bias downward the probability to succeed estimated the following year.

4.3 Identifying the Marginal Effects of Transfers

Until now we have investigated the estimation of the average program impact. While this provides interesting ex-post results on Progresa as it was defined, obtaining marginal effects of transfers would indicate the potential value of modifying the transfer scheme. This marginal impact of transfers is

$$\frac{\partial}{\partial T} E(s_{t+1} \mid \omega_{t+1}, T) \tag{13}$$

where T is the transfer received by the student (the previously defined treatment dummy is $p = 1_{(T>0)}$). To explain the identification method, we derive the following proposition:

Proposition 7 *Assume that there exists a random variable ω'_{t+1} such that the transfer T is $\widetilde{\tau}(g, l, p, \omega'_{t+1})$ and the following assumptions are satisfied:*
The average treatment effect $\frac{\partial}{\partial T} E(s_{t+1} \mid \omega_{t+1}, T)$ does not depend on ω'_{t+1} i.e.

$$\frac{\partial}{\partial \omega'_{t+1}} \left(\frac{\partial}{\partial T} E(s_{t+1} \mid \omega_{t+1}, T)\right) = 0 \qquad \text{(Exclusion Restriction)}$$

The program rule $\widetilde{\tau}(g, l, p, \omega'_{t+1})$ is such that

$$\frac{\partial}{\partial \omega'_{t+1}} \widetilde{\tau}(g, l, p, \omega'_{t+1}) \neq 0 \text{ is known} \qquad \text{(Known Conditionality of Program Rule on Observables)}$$

and does not depend on unobservables $\widetilde{\theta}$

$$\frac{\partial}{\partial \widetilde{\theta}} \{\widetilde{\tau}(g, l, p, \omega'_{t+1})\} = 0 \qquad \text{(Program Rule Independent of Unobservables)}$$

The observed component ω'_{t+1} is independent of unobserved factors $\widetilde{\theta}$ conditionally on ω_{t+1}

$$\omega'_{t+1} \perp\!\!\!\perp \widetilde{\theta} \mid \omega_{t+1} \qquad \text{(IV assumption)}$$

Then $\frac{\partial}{\partial T} E(s_{t+1} \mid \omega_{t+1}, T)$ identifies[12] *$E_{\widetilde{\theta}}[\frac{\partial}{\partial T} E(s_{t+1} \mid \omega_{t+1}, T, \widetilde{\theta})]$.*

[12] Of course only within the range of variation of T in the data observed.

Proof. Proof in Appendix C.6. ■

In Progresa, this identification is provided by the maximum rule as follows.[13] The Progresa rules stipulate that household transfers cannot exceed some given maximum amount of money and impose a proportional adjustment rule for individual benefits. Table 4 gives examples of this proportional adjustment in terms of transfers to be received. The last column shows the transfer due for each child given this adjustment for family A and B which otherwise would get more than the maximum amount allowed while family C does not reach this amount. This monthly amount corresponds to what is lost if a child misses school without justification.

Table 4: Examples of the Maximum Rule

Example of the Maximum Rule in 1997	Progresa Grant	
	Without Adjustment	With Proportional Adjustment
Family A:		
One Boy in Secondary School (1^{st} year)	175	$\frac{175}{605}\times 550 \quad = 159$
One Girl in Secondary School (2^{nd} year)	205	$\frac{205}{605}\times 550 \quad = 186$
One Girl in Secondary School (3^{rd} year)	225	$\frac{225}{605}\times 550 \quad = 205$
Total received by household	605	550
Family B:		
One Girl in Secondary School (1^{st} year)	185	$\frac{185}{585}\times 550 \quad = 173$
One Girl in Secondary School (2^{nd} year)	205	$\frac{205}{585}\times 550 \quad = 193$
One Boy in Secondary School (3^{rd} year)	195	$\frac{195}{585}\times 550 \quad = 184$
Total received by household	585	550
Family C:		
One Girl in Primary School (6^{th} year)	120	120
One Girl in Secondary School (2^{nd} year)	205	205
One Boy in Secondary School (3^{rd} year)	195	195
Total received by household	520	520

Noting as T' the total transfer that the household would receive in absence of this maximum rule and as M_{t+1} the maximum amount of money the household can receive at time $t+1$, the actual transfer received for a child is the known function

$$T = \tau(g,l,p)\min\left\{\frac{M_{t+1}}{T'},1\right\}$$

The assumption needed for identification is that the random variable $\min\left\{\frac{M_{t+1}}{T'},1\right\}$ does not affect the average treatment effect $\frac{\partial}{\partial T}E(s_{t+1}\mid\omega_{t+1},T)$, i.e., that given observables ω_{t+1} the average effect of transfer T on schooling s_{t+1} is constant across values of T'. Concretely, it means that the effect of transfer T on individual schooling can depend on the observable characteristics of a student but that, conditionally on these characteristics ω_{t+1}, there are other observable characteristics that

[13] Without this rule, the value of transfers $T = \tau(g,l,p)$ is conditional only on gender g and grade l which are very likely to be correlated with the individual unobservable components $\widetilde{\theta}$. Then, if transfers do not vary across individuals (conditionally on ω_{t+1}), $\frac{\partial}{\partial T}E(s_{t+1}\mid\omega_{t+1},T)$ is not identifiable and only the average treatment effect $\frac{\partial}{\partial p}E(s_{t+1}\mid\omega_{t+1},p)$ is identified.

affect T' but not the treatment effect. For example, the number of children of the household which generates variation in the individual transfer amount (because some reach the maximum and others not) may have a direct effect on the average treatment effect. However, conditionally on the number of children with for example an equal number of boys and girls, it may be more reasonable to assume that the order of children's gender does not affect the average treatment effect directly while it provides some variation in the amount of transfers received. Table 4 shows examples of families with the same number of children, the same number of boys and girls but for which individual transfers of the girl in the second year of secondary school vary because of this rule. The presence at school of a second year secondary school girl will not bring the same transfer if she belongs to family A, B, or C in the example of Table 4.

We therefore exploit this kind of variation and assume that conditionally on ω_{t+1} (which includes the number of children) the fact that the household reaches the maximum or not is random and uncorrelated with the unobserved characteristics $\widetilde{\theta}$ because it comes mainly from the random distribution of genders within the family. The conditions of identification given by Proposition 7 are then plausible even if not testable.

Identifying how the magnitude of the treatment effect varies with the magnitude of transfers is of great importance for the design of such programs and for their cost-effectiveness analysis (de Janvry and Sadoulet (2006)). An alternative would be to estimate a fully structural model, as Todd and Wolpin (2006) did for child schooling and fertility, in order to simulate the effect of counterfactual values of transfers.

5 Empirical Results and Policy Implications

We now turn to the results on the estimation of the average and marginal effects of Progresa transfers on performance and continuation. Identification of the marginal effect of transfers is based on the 14% of the sample households that are constrained by the maximum benefit rule. Recall that these results are obtained from the first year of operation of the program, hence they reflect an unbiased effect of the program rule, without any selection effect, but they measure very short-term impacts.

5.1 Performance

The probabilities of grade progression during school year 1997-98 are estimated using a logit model with standard errors clustered at the village level. In Table 5, panel (1) reports the average impact of transfers and panel (2) the marginal effect of the transfer amount (with transfers measured in hundreds of pesos). The means of marginal effects denoted $\overline{(\partial\varphi/\partial x)}$ are computed as means of the observation by observation marginal effects for each variable and presented together with coefficient estimates.

Table 5 shows these average effects by school level. Indeed, the program effects ought to be specific to each grade level because of the intrinsic heterogeneity discussed in section 3.1: net effects depend on several "structural" parameters, such as the cost of effort which is higher at higher grades and the "returns" to effort (probability to pass) which can be lower for higher grades, and on the remaining number of years of eligibility. This could potentially lead to a lower impact of the program in the higher grades, even though transfers increase with grade level.

The average treatment effects by gender for primary and secondary school are very different. The means of marginal effects for primary school are positive with a 6% increase in performance for both girls and boys while it is negative in secondary school with means of marginal effects of -22% for girls and -17% for boys. This result confirms the prediction that cash transfers may have a negative impact on learning effort towards the end of eligibility because students want to remain longer in the program.

A full set of control variables included in the regressions show that household size, child male gender, and student's age have negative effects on performance.

Panel (2) of Table 5 reports the estimates with value of transfers included in addition to the Progresa dummy. The estimated marginal effects of transfers are negative in primary school but positive in secondary school, although only significant for male.[14] This contradicts the theory that marginally higher transfers would lead to even more repetition. Estimation of the average treatment effect by gender and grade level of primary and secondary school reported in Table 7 in Appendix D are much more imprecise but consistent with those of Table 5. They show however that the positive marginal effects in secondary school only apply to the first two grades where dynamic incentives may still work but not to the last year, at the end of the program.

[14] Note that the total effect $\theta_1 + \theta_2 \tau$ is always of the sign of the average treatment effect without conditioning on the transfer value for the range of values of effective transfer.

Table 5: Impact on Performance by School Level

$P(l_{t+1} = l_t + 1 \mid s_t = 1)$	(1)				(2)			
	Coeff.		$\overline{(\partial\varphi/\partial x)}$		Coeff.		$\overline{(\partial\varphi/\partial x)}$	
θ_1(grade l,gender g)$.p$								
Primary school, girl	0.552*	(6.76)	0.061*	(7.36)	2.165*	(16.80)	0.194*	(21.26)
Primary school, boy	0.541*	(6.63)	0.060*	(7.13)	2.395*	(18.02)	0.216*	(23.09)
Secondary school, girl	-1.321*	(-5.80)	-0.219*	(-4.73)	-2.753*	(-4.21)	-0.489*	(-4.09)
Secondary school, boy	-1.096*	(-4.56)	-0.172*	(-3.76)	-3.017*	(-4.40)	-0.528*	(-4.65)
θ_2(grade l,gender g)$.\tau(l,g).p$								
Primary school, girl					-2.470*	(-12.45)	-0.287*	(-12.41)
Primary school, boy					-2.897*	(-13.89)	-0.336*	(-13.86)
Secondary school, girl					0.608	(1.71)	0.071	(1.71)
Secondary school, boy					0.925*	(2.37)	0.107*	(2.38)
Covariates X_{t+1}								
Gender (1: boy, 0:girl)	-3.158*	(-8.86)	-0.380*	(-9.03)	-3.856*	(-9.25)	-0.448*	(-9.48)
Household Head Education	-0.009	(-0.30)	-0.001	(-0.30)	-0.019	(-0.62)	-0.002	(-0.62)
Household Size	-0.030*	(-2.46)	-0.004*	(-2.45)	-0.038*	(-2.97)	-0.004*	(-2.95)
Age	-0.308*	(-16.73)	-0.037*	(-16.93)	-0.300*	(-16.10)	-0.035*	(-16.39)
Distance to Sec. School	0.022	(1.51)	0.003	(1.51)	0.021	(1.44)	0.002	(1.44)
Grade×Gender Dummies	Yes				Yes			
State Dummies	Yes				Yes			
Observations	13911				13911			

Note: Transfers τ are in hundreds of pesos, t-stat are in parenthesis, and * means estimate is significantly different from zero at the 5% level. The set of Grade×Gender dummies include the specific grades within primary and secondary levels.

5.2 School Continuation

Equation (7) of school continuation for 1997-98 (in Fall 1998) is estimated and results are shown in Table 6. The probability is estimated using a logit model with standard errors clustered at the village level and means of marginal effects are presented together with coefficient estimates. Control variables show in particular that the household head's education level has a positive effect, age and distance to nearest secondary school a negative effect, while household size and child gender have no significant effect on school continuation.

The average program impact on all students (estimate not shown here) is significantly positive with a 3.5% increase in school continuation, which is consistent with results obtained by Schultz (2004) and Behrman et al. (2005). Table 6 shows that the average treatment effect for primary school is a 3.1% increase in school continuation for girls and boys, and 3.4% for girls and 3.2% for boys in secondary school. Table 8 in Appendix D shows the estimates of grade and gender specific treatments. When significant, there is a positive effect by grade and gender on enrollment in primary and secondary school. The means of marginal effects are around 3-4% for primary school and 3.5% in the 1^{st} year of secondary school but insignificant in the 2^{nd} and 3^{rd} years of secondary school.

Table 6: Impact on Continuation Decision by School Level

$P(s_{t+1}=1 \mid s_t=1)$	(1)				(2)			
	Coeff.		$\overline{(\partial\varphi/\partial x)}$		Coeff.		$\overline{(\partial\varphi/\partial x)}$	
α_1(grade l,gender g).p								
Primary school, girl	0.704*	(3.37)	0.031*	(3.86)	0.774	(1.13)	0.034	(1.29)
Primary school, boy	0.694*	(3.30)	0.031*	(3.78)	1.111	(1.43)	0.046	(1.72)
Secondary school, girl	0.762*	(5.08)	0.034*	(5.79)	-0.981	(-1.19)	-0.058	(-1.01)
Secondary school, boy	0.712*	(4.65)	0.032*	(5.22)	-0.487	(-0.52)	-0.026	(-0.48)
α_2(grade l,gender g).$\tau(l,g)$.p								
Primary school, girl					-0.076	(-0.11)	-0.004	(-0.11)
Primary school, boy					-0.420	(-0.58)	-0.021	(-0.58)
Secondary school, girl					0.876*	(2.14)	0.044*	(2.14)
Secondary school, boy					0.625	(1.30)	0.031	(1.30)
Covariates X_{t+1}								
Gender (1: boy, 0:girl)	1.597	(1.71)	0.079	(1.71)	1.456	(1.39)	0.072	(1.39)
Household Head Education	0.187*	(4.10)	0.009*	(4.12)	0.188*	(4.11)	0.009*	(4.13)
Household Size	0.000	(0.05)	0.000	(0.05)	0.010	(0.58)	0.000	(0.58)
Age	-0.642*	(-20.90)	-0.032*	(-22.09)	-0.643*	(-20.98)	-0.032*	(-22.19)
Distance to Sec. School	-0.101*	(-4.56)	-0.005*	(-4.58)	-0.099*	(-4.50)	-0.005*	(-4.52)
Grade×Gender Dummies	Yes				Yes			
State Dummies	Yes				Yes			
Observations	13894				13894			

Note: Transfers τ are in hundreds of pesos, t-stat are in parenthesis, and * means estimate is significantly different from zero at the 5% level. The set of Grade×Gender dummies include the specific grades within primary and secondary levels.

In Table 6, the marginal effect of transfers is positive for girls in secondary school, with a 100 pesos increase in transfers leading to a 4.4% increase in school continuation. In the grade specific estimates of Table 8, the marginal effect of a one hundred pesos transfer on school continuation is significantly positive for the 1^{st} year of secondary school with means of marginal effects of 4.7% for girls and 5.2% for boys, but not for the last year of school. This suggests that further increasing the transfer for this key transition to secondary school could induce some of the remaining large group of students that quit at the end of primary school to go on with their education.

Todd and Wolpin (2006) use their estimate of a structural model of child schooling and fertility with the Progresa data to engage in counterfactual analysis of other subsidy schemes. While their approach is very different from ours, it is worth noting that our estimates are in the ballpark of their implied marginal effects when predicting changes in completed schooling under counterfactual transfer schemes such as doubling the subsidy. Indeed, from their simulation results (Table 19 in Todd and Wolpin, 2006), we can infer that girls' completed schooling would increase by 6.9% for a doubling of the subsidy. This amounts to a marginal effect of transfers of 6.9% which is larger but in the confidence interval of our marginal effect of transfer estimate for girls in secondary school.

6 Conclusion

We analyzed the effects on school enrollment and performance of Progresa, a conditional cash transfer program in Mexico. We constructed a theoretical framework to analyze the dynamic educational decision and process, including endogeneity and uncertainty of performance in passing grades. This gave predictions for the effect on performance and school continuation of conditional cash transfers for children enrolled at school. To validate the model predictions empirically, we used the randomized experiment set up by Progresa in 506 rural communities. This experiment allowed us to identify the effect of the conditional cash transfers on enrollment and performance at school in the first year of the program, before dynamic selection in passing grades started to bias outcomes. While we must be cautious as we only estimated short-term impacts, we found that Progresa had a positive impact on performance in primary school but a negative impact in secondary school, while having a positive impact on school continuation at all levels as uncovered by Schultz (2004) and Behrman et al. (2005). This empirical finding on performance is a possible consequence of the disincentive effect provided by termination of program benefits after the third year of secondary school.

This paper thus showed the importance of carefully analyzing the behavioral determinants of schooling enrollment and learning when designing financial incentive programs for education in a developing country. Results show that this kind of program modifies the endogenous value function of education and thus the educational behavior of students directly targeted or not by the program. The cash transfers conditional on school attendance proved successful in reducing school drop out but had a perverse negative effect on class repetition, implying that the design of incentive transfers must carefully take into account how the anticipated termination of benefits may affect behavior. It is interesting to observe that, based on lessons learned, Progresa adjusted in subsequent years the design of the program in two ways: first, by extending the cash incentives to the full six years of high school and second by offering a graduation prize consisting in a lump sum transfer to be invested in housing, enterprise startup, or college studies. Through these program improvements, the negative effect of conditional cash transfers on performance that we observed in early years of the program may hopefully have been eliminated.

References

Attanasio O., C. Meghir, and A. Santiago. 2011. "Education Choices in Mexico: Using a Structural Model and a Randomized Experiment to Evaluate Progresa" forthcoming in *Review of Economic Studies*

Behrman J. and Todd P. E. 1999. "Randomness in the experimental samples of Progresa" *Report* *to Progresa, International Food Policy Research Institute, Washington D.C.*

Behrman J., Sengupta P., and Todd P. E. 2005. "Progressing through Progresa: An Impact Assessment of a School Subsidy Experiment in Mexico" *Economic Development and Cultural Change* 54, 1, 237-275

Cameron S. and J. Heckman. 1998. "Life Cycle Schooling and Dynamic Selection Bias: Models and Evidence for Five Cohorts of American Males", *Journal of Political Economy*, 106, 2, 262-333

Cameron S. and J. Heckman. 2001. "The Dynamics of Educational Attainment for Black, Hispanic, and White Males", *Journal of Political Economy*, 109, 3, 455-499

Cameron S. and Taber C.. 2004. "Borrowing Constraints and the Returns to Schooling" *Journal of Political Economy*, 112, 1, 132-182

de Janvry, A. and E. Sadoulet. 2006. "Making Conditional Cash Transfer Programs More Efficient: Designing for Maximum Effect of the Conditionality ", *World Bank Economic Review*, 20, 1, 1-29

De Vreyer P., Lambert S. and Magnac T. 1999. "Educating Children: A Look at Household Behavior in Côte d'Ivoire", *Working Paper INRA LEA n°9905*

Dubois P., A. de Janvry, and E. Sadoulet. 2007. "Effects on School Enrollment and Performance of a Conditional Cash Transfer Program in Mexico" CEPR DP 6069

Eckstein Z. and K. Wolpin. 1999. "Why Youths Drop out of High School: The Impact of Preferences, Opportunities and Abilities", *Econometrica,* 65, 1295-1340

Ham J. and Lalonde R. 1996. "The Effect of Sample Selection and Initial Conditions in Duration Models: Evidence from Experimental Data on Training", *Econometrica*, 64, 1, 175-206

Heckman, J. and B. Singer. 1984. "The Identifiability of the Proportional Hazard Model." *The Review of Economic Studies* 51(2): 231-241.

Heckman, J. and Vytlacil E. 1999. "Local Instrumental Variables and Latent Variable Models for Identifying and Bounding Treatment Effects," Proceedings of the National Academy of Sciences, 96, 4730-4734

Heckman, J. and Vytlacil E. 2001. "Local Instrumental Variables" in Nonlinear Statistical Interference: Essays in Honor of Takeshi Amemiya, edited by C. Hsiao, K. Morimune, and J. Powell, (Cambridge: Cambridge University Press)

Keane M. P. and Wolpin K. I. 1997. "The Career Decisions of Young Men", *Journal of Political Economy*, 105, 3, 473-522

Keane M. P. and Wolpin K. I. 2001. "The Effect of Parental Transfers and Borrowing Constraints on Educational Attainment", *International Economic Review*, 42, 4, 1051-1102

Magnac T. and D. Thesmar. 2002a. "Analyse économique des politiques éducatives: l'augmentation de la scolarisation en France de 1982 à 1993", *Annales d'Economie et de Statistiques*, 66, 1-35

Magnac T. and D. Thesmar. 2002b. "Identifying Dynamic Discrete Decision Processes", *Econometrica*, 70, 2, 801-816

Manski C. 1985. "Semiparametric Analysis of Discrete Response: Asymptotic Properties of the Maximum Score Estimator", *Journal of Econometrics*, 27, 3, 313-334

Manski C. 1988. "Identification of Binary Response Models", *Journal of the American Statistical Association*, 83, 403, 729-738

Progresa. 1999. "Evaluacion de Resultados del Programa de Educacion, Salud y Alimentacion - Primeros Avances" Report Edited by Progresa, Mexico City.

Progresa. 2000. "Reglas Generales para la Operacion del Programa de Educacion, Salud y Alimentacion" Report Edited by Progresa, Mexico City.

Rosenzweig M. and K. Wolpin. 1995. "Sisters, Siblings, and Mothers: The Effect of Teen-Age Childbearing on Birth Outcomes in a Dynamic Family Context" *Econometrica*, 63, 2, 303-326

Schultz T. P. 2000. "The Impact of Progresa on School Enrollments" *Final Report, International Food Policy Research Institute, Washington D.C.*

Schultz T. P. 2004. "School subsidies for the poor: evaluating the Mexican Progresa poverty program" *Journal of Development Economics,* 74, 1, 199-250

Skoufias E., Davis B. and Behrman J. R. 1999. "An evaluation of the selection of beneficiary households in the Education, Health and Nutrition Program (Progresa) of Mexico" *Report to Progresa, International Food Policy Research Institute, Washington D.C.*

Stokey N. and Lucas R. 1989. *Recursive Methods in Economic Dynamics*, Cambridge, Mass.: Harvard University Press

Taber C. 2001. "The Rising College Premium in the Eighties: Return to College or Return to Unobserved Ability?" *Review of Economic Studies*, 68, 3, 665-691.

Todd, P. E., and K. Wolpin. 2006. "Using a Social Experiment to Validate a Dynamic Behavioral Model of Child Schooling and Fertility: Assessing the Impact of a School Subsidy Program in Mexico" *American Economic Review* 96, 5, 1384-1417.

A Data

The data used were provided by Progresa. To construct our variables and sample we used all the available relevant information from the following data sets: ENCASEH97 (Encuesta de Características Socioeconomicas de los Hogares), ENCEL (Encuesta de Evaluacion) of March 1998 and October 1998.

The schooling variable used in the analysis corresponds to the question on whether the child is currently going to school which means both enrollment and non permanent absenteeism.

The variables on grades correspond to the question on what is the last grade completed by the child. It is assumed that he or she is then entitled to enroll at the upper grade. We use the intermediate evaluation survey of March 1998 to check the consistency of the data, correct the obviously erroneous answers, and to complete non responses that sometimes happen in a given survey.

B Increasing Value function

Corollary 8 *The value function ϕ is increasing in l if individuals are sufficiently patient ($\beta \geq 1/2$) and transfers are increasing with grade (or not too much decreasing i.e. transfers not too large when the program ends) and not too large compared to the potential wage (conditions likely to be true in the case of Progresa).*

Proof. Simplifying notations by avoiding indices g and p when there is no ambiguity, we know that ϕ exists, is unique and is the fixed point solution of T where $\forall l : T(\phi)(l) = \max\{w(g,l) + \beta\phi(l), \max_{e\geq 0}\{\tau(l,g,p) - c(e) + \beta E[\phi(l'(l)) \mid s=1]\}\}$ *s.t.* $l'(l) = l+1$ if $s=1$ and $\pi(l,e) \geq \varepsilon$ and $l' = l$ otherwise. T being a contraction mapping, the fixed point solution will be increasing in l if ϕ increasing in l implies $T(\phi)$ increasing in l.

$T(\phi)(l+1) - T(\phi)(l)$

$= \max\{w(g,l+1) + \beta\phi(l+1), \max_e\{\tau(l+1,g,p) - c(e) + \beta E[\phi(l'(l+1)) \mid s=1]\}\}$

$- \max\{w(g,l) + \beta\phi(l), \max_e\{\tau(l,g,p) - c(e) + \beta E[\phi(l'(l)) \mid s=1]\}\}$ where $l'(l) = l+1$ if $s=1$ and $\pi(l,e) \geq \varepsilon$ and $l' = l$ otherwise.

So $T(\phi)(l+1) - T(\phi)(l)$ is one of the four following values:

a) $w(g,l+1) - w(g,l) + \beta[\phi(l+1) - \phi(l)] \geq 0$ because w and ϕ are increasing in l.

b) $\tau(l+1,g,p) - \tau(l,g,p) + \beta\max_e\{\{E[\phi(l'(l+1)) \mid s=1] - c(e)\} - \beta\max_e\{E[\phi(l'(l)) \mid s = 1] - c(e)\}$ is likely to be positive if $\tau(l+1,g,p) - \tau(l,g,p)$ is positive or not too large compared to the discounted marginal value of a higher education degree.

c) $w(g,l+1) - \tau(l,g,p) + \beta\phi(l+1) - \max_e\{\beta E[\phi(l'(l)) \mid s=1] - c(e)\} \geq 0$ if $\tau(l,g,p) \leq w(g,l+1)$ (the wage is sufficiently high compared to the transfer) because $E[\phi(l') \mid s=1] < \phi(l+1)$ which implies that $\beta\phi(l+1) - \max_e\{\beta E[\phi(l') \mid s=1] - c(e)\} > 0$.

d) $\tau(l+1,g,p) + \max_e\{\beta E[\phi(l'(l+1)) \mid s=1] - c(e)\} + \beta\phi(l) - w(g,l) > 0$ because obviously $\phi(l) > \frac{w(g,l)}{1-\beta}$ implying that $\beta\phi(l) > \frac{\beta}{1-\beta}w(g,l) \geq w(g,l)$ if $\beta \geq 1/2$ and $\max_e\{\beta E[\phi(l'(l+1)) \mid s=1] - c(e)\} \geq 0$. ■

C Proofs

C.1 Proof of Proposition 1

Noting $\phi(.,g,p) = \phi(.)$, let's first define an operator $T_{g,p}$ transforming $\phi(.)$ in $T_{g,p}(\phi(.))$ by

$$\begin{aligned} \forall l &: T_{g,p}(\phi)(l) = \max\{w(g,l) + \beta\phi(l), \max_{e\geq 0}\{\tau(l,g,p) - c(e) + \beta E[\phi(l') \mid s=1]\}\} \\ s.t.\ l' &= l+1 \text{ if } s=1 \text{ and } \pi(l,e) \geq \varepsilon \text{ and } l'=l \text{ otherwise.} \end{aligned}$$

$\phi(.)$ is the fixed point (if any) of the operator $T_{g,p}(.)$. If $T_{g,p}(.)$ is a contraction mapping, then its fixed point exists and is unique (see Stokey and Lucas, 1989). Using Blackwell sufficiency theorem, we just need to show that $T_{g,p}$ verifies the monotonicity and discounting properties. Let $\phi, \widetilde{\phi} \in C(\Re_+, \Re)$ such that $\forall l$, $\phi(l) \leq \widetilde{\phi}(l)$ then it is straightforward to check that $\forall l$, $T_{g,p}\phi(l) \leq T_{g,p}\widetilde{\phi}(l)$ (monotonicity property). Moreover $\forall l$, $T_{g,p}(\phi+\gamma)(l) \leq T_{g,p}(\phi)(l) + \beta\gamma$ because it is straightforward to check that $T_{g,p}(\phi+\gamma)(l) = T_{g,p}(\phi)(l) + \beta\gamma$. ϕ is the fixed point of T: $T_{g,p}(\phi) = \phi$.

C.2 Proof of Proposition 2

Conditional on schooling, the learning effort is chosen to maximize $\tau(l,g,p) - c(e) + \beta E[\phi_{g,p}(l') \mid s=1,e]$. This program is always concave in e because $E[\phi_{g,p}(l') \mid s=1,e] = P(l'=l+1 \mid s=1,e)\phi(l+1,g,p) + P(l'=l \mid s=1,e)\phi(l,g,p)$

$= \phi(l,g,p) + P(l'=l+1 \mid s=1,e)[\phi(l+1,g,p) - \phi(l,g,p)]$ and $P(l'=l+1 \mid s=1,e)$ is increasing concave in e. Then, the learning effort will satisfy the first order condition $\beta\frac{\partial}{\partial e}E[\phi_{g,p}(l) \mid s=1] = c'(e)$. Since $P(l'=l+1 \mid s=1,e) = F \circ \pi(l,e^*)$ and when $\phi(l+1,g,p) > \phi(l,g,p)$ the first order condition equation determining e^* is

$$\beta[\phi(l+1,g,p) - \phi(l,g,p)]f \circ \pi(l,e^*)\frac{\partial\pi}{\partial e}(l,e^*) = c'(e^*)$$

When $\phi(l+1,g,p) - \phi(l,g,p) < 0$ then $e^* = 0$. Since f is decreasing, $\pi(l,.)$ is increasing concave in e, and $c(.)$ increasing convex, we can use the implicit function theorem to find some properties of e^*. If $\frac{\partial\pi}{\partial l} = 0$, e^* has the same directions of variation in l than $\phi(l+1,g,p) - \phi(l,g,p)$. This implies that e^* is increasing in l if $\phi(l,g,p)$ is convex in l and decreasing in l if $\phi(l,g,p)$ is concave in l. If $\frac{\partial\pi}{\partial l} \neq 0$, then e^* can be either increasing or decreasing in l, according to the properties of $\frac{\partial^2\pi}{\partial e\partial l}$ and $\phi(l+1,g,p) - \phi(l,g,p)$.

C.3 Proof of Proposition 4

Remind that $v(l_t, g, p) = V(l_t, g, p, 1) - V(l_t, g, p, 0)$ where

$$\begin{aligned} V(l_t, g, p, 1) &= \max_{e_t}\{\tau(l_t, g, p) - c(e_t) + \beta E[\phi(l_{t+1}, g, p) \mid s_t = 1]\} \\ V(l_t, g, p, 0) &= w(g, l_t) + \beta\phi(l_t, g, p) \end{aligned}$$

Using the previous definitions and $\tau(l_t, g, p) = 1_{\{p=1\}}\tau(l_t, g, 1)$ since $\tau(l_t, g, 0) = 0$,

$$\begin{aligned} v(l_t, g, p) &= 1_{\{p=1\}}\tau(l_t, g, 1) + \max_{e_t}\{-c(e_t) + \beta E[\phi(l_{t+1}, g, p) \mid s_t = 1]\} \\ &\quad -w(g, l_t) - \beta\phi(l_t, g, p) \end{aligned}$$

Then

$$\begin{aligned} \Delta v(l_t, g, p) &= \tau(l_t, g, 1) - \beta\Delta\phi(l_t, g, p) \\ &\quad +\Delta\left[\max_{e_t}\{-c(e_t) + \beta E[\phi(l_{t+1}, g, p) \mid s_t = 1]\}\right] \end{aligned}$$

Remark that

$$E[\phi(l_{t+1}, g, p) \mid s_t = 1] = \phi(l_t + 1, g, p) P(l_{t+1} = l_t{+}1 \mid s_t = 1, e_t) + \phi(l_t, g, p) P(l_{t+1} = l_t \mid s_t = 1, e_t)$$

where $P(l_{t+1} = l_t + 1 \mid s_t = 1, e_t)$ is the probability of success given effort e_t. However, we cannot use the envelope theorem as if p was continuous, thus we have

$$\begin{aligned} \Delta E[\phi(l_{t+1}, g, p) \quad &\mid \quad s_t = 1] = P(l_{t+1} = l_t + 1 \mid s_t = 1)\Delta\phi(l_t + 1, g, p) \\ +\phi(l_t + 1, g, p)\Delta P(l_{t+1} &= l_t + 1 \mid s_t = 1) \\ +P(l_{t+1} &= l_t \mid s_t = 1)\Delta\phi(l_t, g, p) + \phi(l_t, g, p)\Delta P(l_{t+1} = l_t \mid s_t = 1) \end{aligned}$$

which leads to

$$\begin{aligned} \Delta v(l_t, g, p) &= \tau(l_t, g, 1) - \beta\Delta\phi(l_t, g, p) - \Delta c(e(p)) \\ +\beta P(l_{t+1} &= l_t + 1 \mid s_t = 1)\Delta\phi(l_t + 1, g, p) + \beta\phi(l_t + 1, g, p)\Delta P(l_{t+1} = l_t + 1 \mid s_t = 1) \\ +\beta P(l_{t+1} &= l_t \mid s_t = 1)\Delta\phi(l_t, g, p) + \beta\phi(l_t, g, p)\Delta P(l_{t+1} = l_t \mid s_t = 1) \end{aligned}$$

As the optimal effort is an increasing function of the probability of success by inverting the increasing function $P(l_{t+1} = l_t + 1 \mid s_t = 1, e_t)$, we can write $c(e) = h(P(l_{t+1} = l_t + 1 \mid s_t = 1))$ where h is an increasing function

$$\begin{aligned} \Delta v(l_t, g, p) &= \tau(l_t, g, 1) + \beta P(l_{t+1} = l_t + 1 \mid s_t = 1)\Delta\left[\phi(l_t + 1, g, p) - \phi(l_t, g, p)\right] \\ -\Delta h(P(l_{t+1} &= l_t + 1 \mid s_t = 1)) + \beta\left[\phi(l_t + 1, g, p) - \phi(l_t, g, p)\right]\Delta P(l_{t+1} = l_t + 1 \mid s_t = 1) \end{aligned}$$

which gives the proposition.

C.4 Proof of Proposition 5

It comes from the fact that the conditional distribution of $\widetilde{\theta}$ on ω_{t+1} does not depend on p. Using the law of iterated expectations, a simple differentiation of the expectation proves it[15]:

$$\begin{aligned}
\Delta E(s_{t+1} \mid \omega_{t+1}, p) &= \Delta \iint s_{t+1} d\lambda(s_{t+1}, \widetilde{\theta} \mid \omega_{t+1}, p) \\
&= \Delta \iint s_{t+1} d\lambda(s_{t+1} \mid \omega_{t+1}, p, \widetilde{\theta}) d\lambda(\widetilde{\theta} \mid \omega_{t+1}, p) \\
&= \int [\Delta \int s_{t+1} d\lambda(s_{t+1} \mid \omega_{t+1}, p, \widetilde{\theta})] d\lambda(\widetilde{\theta} \mid \omega_{t+1}) = E_{\widetilde{\theta}}[\Delta\psi(\omega_{t+1}, p, \widetilde{\theta})]
\end{aligned}$$

because (9) implies that $d\lambda(\widetilde{\theta} \mid \omega_{t+1}, p) = d\lambda(\widetilde{\theta} \mid \omega_{t+1}, 1) = d\lambda(\widetilde{\theta} \mid \omega_{t+1}, 0) = d\lambda(\widetilde{\theta} \mid \omega_{t+1})$.

C.5 Proof of Proposition 6

Defining biases equal to $B(\omega_{t+1}, 1) = \frac{\partial}{\partial \omega_{t+1}}[E_{\widetilde{\theta}} E(s_{t+1} \mid \omega_{t+1}, 1, \widetilde{\theta})] - E_{\widetilde{\theta}}[\frac{\partial}{\partial \omega_{t+1}} E(s_{t+1} \mid \omega_{t+1}, 1, \widetilde{\theta})]$ for the treatment population and to $B(\omega_{t+1}, 0) = \frac{\partial}{\partial \omega_{t+1}}[E_{\widetilde{\theta}} E(s_{t+1} \mid \omega_{t+1}, 0, \widetilde{\theta})] - E_{\widetilde{\theta}}[\frac{\partial}{\partial \omega_{t+1}} E(s_{t+1} \mid \omega_{t+1}, 0, \widetilde{\theta})]$ for the control population.

The difference of biases is

$$\begin{aligned}
\Delta B(\omega_{t+1}, p) &= \Delta\{\frac{\partial}{\partial \omega_{t+1}}[E_{\widetilde{\theta}} \psi(\omega_{t+1}, p, \widetilde{\theta})] - E_{\widetilde{\theta}}[\frac{\partial}{\partial \omega_{t+1}} \psi(\omega_{t+1}, p, \widetilde{\theta})]\} \\
&= \frac{\partial}{\partial \omega_{t+1}} E_{\widetilde{\theta}}[\Delta\psi(\omega_{t+1}, p, \widetilde{\theta})] - E_{\widetilde{\theta}}[\Delta \frac{\partial}{\partial \omega_{t+1}} \psi(\omega_{t+1}, p, \widetilde{\theta})]
\end{aligned}$$

because $d\lambda(\widetilde{\theta} \mid \omega_{t+1}, p) = d\lambda(\widetilde{\theta} \mid \omega_{t+1})$. However

$$\frac{\partial}{\partial \omega_{t+1}} E_{\widetilde{\theta}}[\Delta\psi(\omega_{t+1}, p, \widetilde{\theta})] = \frac{\partial}{\partial \omega_{t+1}} \int [\Delta \int s_{t+1} d\lambda(s_{t+1} \mid \omega_{t+1}, p, \widetilde{\theta})] d\lambda(\widetilde{\theta} \mid \omega_{t+1})$$

$$= \int [\Delta \int s_{t+1} d\lambda(s_{t+1} \mid \omega_{t+1}, p, \widetilde{\theta})] \frac{\partial}{\partial \omega_{t+1}} d\lambda(\widetilde{\theta} \mid \omega_{t+1}) + \int \frac{\partial}{\partial \omega_{t+1}} [\Delta \int s_{t+1} d\lambda(s_{t+1} \mid \omega_{t+1}, p, \widetilde{\theta})] d\lambda(\widetilde{\theta} \mid \omega_{t+1})$$

$$= \int [\Delta \int s_{t+1} d\lambda(s_{t+1} \mid \omega_{t+1}, p, \widetilde{\theta})] \frac{\partial}{\partial \omega_{t+1}} d\lambda(\widetilde{\theta} \mid \omega_{t+1}) + E_{\widetilde{\theta}}[\Delta \frac{\partial}{\partial \omega_{t+1}} \psi(\omega_{t+1}, p, \widetilde{\theta})]$$

Then $\Delta B(\omega_{t+1}, p) = 0$ if and only if $\int [\Delta \int s_{t+1} d\lambda(s_{t+1} \mid \omega_{t+1}, p, \widetilde{\theta})] \frac{\partial}{\partial \omega_{t+1}} d\lambda(\widetilde{\theta} \mid \omega_{t+1}) = 0$. This is not always true since in general $\frac{\partial}{\partial \omega_{t+1}} d\lambda(\widetilde{\theta} \mid \omega_{t+1}) \neq 0$.

Noting ω_{t+1}^k some component of ω_{t+1}, the average change according to ω_{t+1}^k in the impact of p, $E_{\widetilde{\theta}}[\Delta \frac{\partial}{\partial \omega_{t+1}^k} \psi(\omega_{t+1}, p, \widetilde{\theta})]$, will be identified through the estimation of $\frac{\partial}{\partial \omega_{t+1}^k} E_{\widetilde{\theta}}[\Delta\psi(\omega_{t+1}, p, \widetilde{\theta})]$ if and only if one of the following conditions is satisfied:

$$\frac{\partial}{\partial \omega_{t+1}^k} d\lambda(\widetilde{\theta} \mid \omega_{t+1}) = 0$$

[15] The notation $\lambda(\mu \mid \nu)$ always means the cumulative distribution of μ conditional on ν.

or

$$\int \Delta E(s_{t+1} \mid \omega_{t+1}, p, \widetilde{\theta}) \frac{\partial}{\partial \omega_{t+1}^k} d\lambda(\widetilde{\theta} \mid \omega_{t+1}) = 0$$

$$\Leftrightarrow \int E(s_{t+1} \mid \omega_{t+1}, 1, \widetilde{\theta}) \frac{\partial}{\partial \omega_{t+1}^k} d\lambda(\widetilde{\theta} \mid \omega_{t+1}) = \int E(s_{t+1} \mid \omega_{t+1}, 0, \widetilde{\theta}) \frac{\partial}{\partial \omega_{t+1}^k} d\lambda(\widetilde{\theta} \mid \omega_{t+1})$$

Then, we just need to use Proposition 5 to complete the proof.

C.6 Proof of Proposition 7

We just need to prove that $\frac{\partial}{\partial T} E\left(s_{t+1} \mid \omega_{t+1}, T\right) = E_{\widetilde{\theta}}[\frac{\partial}{\partial T} E(s_{t+1} \mid \omega_{t+1}, T, \widetilde{\theta})]$. We have $\frac{\partial}{\partial T} E\left(s_{t+1} \mid \omega_{t+1}, T\right) = \frac{\partial}{\partial T}[E_{\widetilde{\theta}} \psi(\omega_{t+1}, T, \widetilde{\theta})]$ where $E(s_{t+1} \mid \omega_{t+1}, T, \widetilde{\theta}) = \psi(\omega_{t+1}, T, \widetilde{\theta})$. With $T = \widetilde{\tau}(g, l, p, \omega'_{t+1})$ and $\frac{\partial}{\partial \omega'_{t+1}} \widetilde{\tau}(g, l, p, \omega'_{t+1})$ known (by the program rule),

$$\frac{\partial}{\partial T} \psi(\omega_{t+1}, T, \widetilde{\theta}) = \frac{\frac{\partial}{\partial \omega'_{t+1}} \psi(\omega_{t+1}, \widetilde{\tau}(g, l, p, \omega'_{t+1}), \widetilde{\theta})}{\frac{\partial}{\partial \omega'_{t+1}} \widetilde{\tau}(g, l, p, \omega'_{t+1})}.$$

Thus, we can write

$$\begin{aligned}
E_{\widetilde{\theta}}[\frac{\partial}{\partial T} \psi(\omega_{t+1}, T, \widetilde{\theta})] &= \frac{E_{\widetilde{\theta}}[\frac{\partial}{\partial \omega'_{t+1}} \psi(\omega_{t+1}, \widetilde{\tau}(g, l, p, \omega'_{t+1}), \widetilde{\theta})]}{\frac{\partial}{\partial \omega^k_{t+1}} \widetilde{\tau}(g, l, p, \omega'_{t+1})} \text{ because } \frac{\partial}{\partial \widetilde{\theta}} \{\frac{\partial}{\partial \omega'_{t+1}} \widetilde{\tau}(g, l, p, \omega'_{t+1})\} = 0 \\
&= \frac{\frac{\partial}{\partial \omega'_{t+1}} E_{\widetilde{\theta}}[\psi(\omega_{t+1}, \widetilde{\tau}(g, l, p, \omega'_{t+1}), \widetilde{\theta})]}{\frac{\partial}{\partial \omega'_{t+1}} \widetilde{\tau}(g, l, p, \omega'_{t+1})} \text{ because } \widetilde{\theta} \perp\!\!\!\perp \ \omega'_{t+1} \mid \omega_{t+1} \\
&= \frac{\frac{\partial}{\partial \omega'_{t+1}} E\left(s_{t+1} \mid \omega_{t+1}, \widetilde{\tau}(g, l, p, \omega'_{t+1})\right)}{\frac{\partial}{\partial \omega'_{t+1}} \widetilde{\tau}(g, l, p, \omega'_{t+1})}
\end{aligned}$$

which proves the proposition since

$$\frac{\partial}{\partial T} E\left(s_{t+1} \mid \omega_{t+1}, T\right) = \frac{\frac{\partial}{\partial \omega'_{t+1}} E(s_{t+1} \mid \omega_{t+1}, \widetilde{\tau}(g, l, p, \omega'_{t+1}))}{\frac{\partial}{\partial \omega'_{t+1}} \widetilde{\tau}(g, l, p, \omega'_{t+1})}.$$

D Additional Tables

Table 7: Impact on Performance by Grade

$P(l_{t+1} = l_t + 1 \mid s_t = 1)$	(1)				(2)			
(t-stat), (*: 5% significance)	Coeff.		$\overline{(\partial\varphi/\partial x)}$		Coeff.		$\overline{(\partial\varphi/\partial x)}$	
θ_1(grade l,gender g).p								
P3, girl	0.418*	(2.99)	0.052*	(3.31)	0.071	(0.08)	0.009	(0.09)
P3, boy	0.336*	(2.46)	0.043*	(2.66)	-0.498	(-0.64)	-0.074	(-0.58)
P4, girl	-0.320	(-1.77)	-0.048	(-1.66)	0.020	(0.02)	0.003	(0.02)
P4, boy	-0.234	(-1.28)	-0.034	(-1.22)	-1.342	(-1.50)	-0.228	(-1.29)
P5, girl	-0.056	(-0.25)	-0.008	(-0.25)	-1.649	(-1.48)	-0.290	(-1.29)
P5, boy	-0.180	(-0.83)	-0.026	(-0.80)	-1.527	(-1.74)	-0.265	(-1.51)
P6, girl	-0.177	(-0.69)	-0.025	(-0.67)	-2.357*	(-2.82)	-0.431*	(-2.75)
P6, boy	-0.312	(-1.23)	-0.046	(-1.15)	0.036	(0.03)	0.005	(0.04)
S1, girl	-1.989*	(-5.50)	-0.378*	(-5.08)	-2.467*	(-3.46)	-0.460*	(-3.45)
S1, boy	-1.617*	(-4.41)	-0.297*	(-3.88)	-3.226*	(-4.25)	-0.582*	(-5.52)
S2, girl	-0.333	(-0.66)	-0.050	(-0.62)	-2.682	(-1.71)	-0.501	(-1.81)
S2, boy	-0.296	(-0.62)	-0.044	(-0.58)	-3.282*	(-2.38)	-0.591*	(-3.16)
S3, girl	-1.091*	(-2.21)	-0.190	(-1.89)	-1.258	(-0.94)	-0.216	(-0.79)
S3, boy	-1.470*	(-2.72)	-0.269*	(-2.34)	-1.513	(-1.26)	-0.268	(-1.07)
θ_2(grade l,gender g).$\tau(l,g)$								
P3, girl					0.602	(0.42)	0.080	(0.42)
P3, boy					1.451	(1.09)	0.193	(1.09)
P4, girl					-0.517	(-0.37)	-0.069	(-0.37)
P4, boy					1.646	(1.23)	0.219	(1.23)
P5, girl					1.859	(1.43)	0.247	(1.43)
P5, boy					1.583	(1.55)	0.210	(1.55)
P6, girl					1.983*	(2.63)	0.263*	(2.63)
P6, boy					-0.316	(-0.36)	-0.042	(-0.36)
S1, girl					0.284	(0.74)	0.038	(0.74)
S1, boy					1.002*	(2.22)	0.133*	(2.22)
S2, girl					1.290	(1.50)	0.171	(1.51)
S2, boy					1.778*	(2.15)	0.236*	(2.15)
S3, girl					0.085	(0.14)	0.011	(0.14)
S3, boy					0.033	(0.05)	0.004	(0.05)
Gender (1: boy, 0:girl)	-3.718*	(-7.13)	-0.514*	(-7.19)	0.482	(0.74)	0.064	(0.74)
Household Head Education	-0.021	(-0.66)	-0.003	(-0.66)	-0.022	(-0.71)	-0.003	(-0.71)
Household Size	-0.031*	(-2.49)	-0.004*	(-2.49)	-0.018	(-1.36)	-0.002	(-1.36)
Age	-0.308*	(-15.99)	-0.043*	(-16.52)	-0.311*	(-16.01)	-0.041*	(-16.49)
Distance to Sec. School	0.017	(1.14)	0.002	(1.14)	0.018	(1.19)	0.002	(1.19)
Grade×Gender Dummies	Yes				Yes			
State Dummies	Yes				Yes			
Observations	13911				13911			

Note: All additive gender-grade dummies as well as state dummies are not shown but included.

Table 8: Impact on Continuation Decision by Grade

$P(s_{t+1}=1 \mid s_t=1)$	(1)				(2)			
(t-stat), (*: 5% significance)	Coeff.		$\overline{(\partial\varphi/\partial x)}$		Coeff.		$\overline{(\partial\varphi/\partial x)}$	
α_1(grade l,gender g).p								
P2, girl	-0.322	(-0.35)	-0.017	(-0.33)	-0.319	(-0.35)	-0.017	(-0.33)
P3, girl	1.104*	(2.63)	0.043*	(3.54)	-1.773	(-0.40)	-0.129	(-0.29)
P3, boy	0.536	(1.09)	0.024	(1.25)	0.542	(0.08)	0.024	(0.09)
P4, girl	1.203*	(2.36)	0.045*	(3.35)	-3.860	(-1.01)	-0.410	(-0.66)
P4, boy	0.901*	(2.24)	0.037*	(2.84)	1.668	(0.36)	0.057	(0.56)
P5, girl	0.630	(1.77)	0.027*	(2.06)	-2.767	(-1.20)	-0.241	(-0.80)
P5, boy	0.837*	(2.15)	0.034*	(2.66)	3.699	(0.84)	0.088	(1.64)
P6, girl	0.390	(1.00)	0.018	(1.09)	5.158	(0.92)	0.105*	(2.31)
P6, boy	0.459	(1.32)	0.021	(1.47)	-2.903	(-0.89)	-0.258	(-0.58)
S1, girl	0.857*	(5.25)	0.037*	(6.18)	-1.004	(-1.14)	-0.060	(-0.95)
S1, boy	0.823*	(4.94)	0.035*	(5.77)	-1.159	(-1.16)	-0.071	(-0.94)
S2, girl	0.133	(0.29)	0.006	(0.29)	-3.686*	(-2.28)	-0.383	(-1.46)
S2, boy	0.629	(1.25)	0.027	(1.47)	-1.110	(-0.49)	-0.070	(-0.40)
S3, girl	0.280	(0.39)	0.013	(0.42)	2.792	(0.88)	0.072*	(2.25)
S3, boy	-0.174	(-0.33)	-0.009	(-0.31)	4.307	(1.02)	0.088*	(3.62)
α_2(grade l,gender g).$\tau(l,g)$								
P3, girl					4.236	(0.64)	0.210	(0.64)
P4, girl					6.550	(1.30)	0.325	(1.30)
P4, boy					-0.970	(-0.16)	-0.048	(-0.16)
P5, girl					3.512	(1.48)	0.174	(1.48)
P5, boy					-2.905	(-0.66)	-0.144	(-0.66)
P6, girl					-3.619	(-0.86)	-0.179	(-0.86)
P6, boy					2.577	(1.04)	0.128	(1.03)
S1, girl					0.947*	(2.13)	0.047*	(2.13)
S1, boy					1.048*	(2.01)	0.052*	(2.01)
S2, girl					1.840*	(2.36)	0.091*	(2.36)
S2, boy					0.894	(0.77)	0.044	(0.77)
S3, girl					-1.081	(-0.84)	-0.054	(-0.84)
S3, boy					-2.119	(-1.08)	-0.105	(-1.08)
Covariates X_{t+1}								
Gender (1: boy, 0:girl)	0.137	(0.63)	0.007	(0.63)	-1.382	(-1.21)	-0.069	(-1.20)
Household Head Education	0.188*	(4.12)	0.009*	(4.14)	0.188*	(4.12)	0.009*	(4.14)
Household Size	0.000	(0.03)	0.000	(0.03)	0.015	(0.83)	0.000	(0.83)
Age	-0.643*	(-20.78)	-0.032*	(-22.06)	-0.650*	(-21.03)	-0.032*	(-22.11)
Distance to Sec. School	-0.101*	(-4.59)	-0.005*	(-4.60)	-0.098*	(-4.46)	-0.005*	(-4.48)
Grade×Gender Dummies	Yes				Yes			
State Dummies	Yes				Yes			
Observations	13894				13894			

Note: All additive gender-grade dummies as well as state dummies are not shown but included.