THE WORLD BANK ECONOMIC REVIEW, VOL. 13, NO. 3: 493-508

Student Outcomes in Philippine Elementary Schools: An Evaluation of Four Experiments

Jee-Peng Tan, Julia Lane, and Gerard Lassibille

Policymakers in most developing countries are concerned about high dropout rates and poor student learning in primary education. The government of the Philippines initiated the Dropout Intervention Program in 1990–92 as part of its effort to address these issues. Under this program, four experimental interventions were randomly assigned to 20 schools in selected low-income areas. Pre- and post-intervention data were collected from these schools, as well as from 10 control schools, in order to evaluate the program's impact on dropout behavior and student learning. The economic justification for replication appears to be strongest for the interventions that provided teachers with learning materials, which helped them to pace lessons according to students' differing abilities, and that initiated parent-teacher partnerships, which involved parents in the schooling of their children. The justification was weakest for the school feeding intervention. In addition to the results specific to the Philippines, this research demonstrates the feasibility of monitoring and evaluating interventions in the education sector in other developing countries, including the use of randomized control designs.

Most developing countries now recognize that investing in education, particularly primary education, provides an essential bedrock for economic and social development. In past decades governments emphasized expanding enrollment, but as coverage rose, the problems of low completion rates and inadequate student learning came to the fore (see Lockheed and Verspoor 1991 for a comprehensive treatment of these issues). Policymakers need information on the costs as well as on the impact of different methods of improving schooling outcomes. Unfortunately, however, the literature on the quantitative relationship between inputs and outcomes in education is sparse, and most developing countries have only a nascent capacity to conduct their own context-specific research and evaluation.[1]

1. See Harbison and Hanushek (1992) for a summary of results from 96 studies on the relationship between school inputs and learning based on data from developing countries. Quantitative studies on the relationship between school inputs and dropout behavior and between school inputs and grade repetition are much more rare. Recent examples include Hanushek and Lavy (1994), Gomes-Neto and Hanushek (1994), and Chuard and Mingat (1996).

Jee-Peng Tan is with the Human Development Network at the World Bank, Julia Lane is with the Urban Institute, and Gerard Lassibille is with the Centre National de la Recherche Scientifique in Dijon, France. The authors' e-mail addresses are jtan@worldbank.org, jlane@ui.urban.org, and gerard.lassibille@u-bourgogne.fr.
The authors thank Ms. Lidi Santos and her staff at the Bureau of Elementary Education of the Philippines Department of Education, Culture, and Sports for facilitating access to the data, the World Bank's Research Department for partial financial support, Alain Mingat for helpful discussions on the analytical method, and three anonymous referees for comments and suggestions on earlier drafts.

This article documents an evaluation effort in the Philippines intended to guide policymaking in primary education. Almost all children in the country enter first grade, but not all of them reach the end of the primary school cycle. Data for the early 1990s suggest that noncompleters represent about 25 percent of each entering cohort of first graders. Further, many children leave school having learned only a fraction of the primary school curriculum (Miguel 1993). Achievement tests administered in the 1991 Household and School Matching Survey, for example, show that pupils in grades two to six mastered less than half of the curriculum they were taught (Tan, Lane, and Coustère 1997).

As part of its strategy to improve primary education, the Philippine government implemented the Dropout Intervention Program (DIP) in the context of a World Bank–financed elementary education project. The program comprised four experimental interventions. Each of these was implemented in five schools during the 1991–92 academic year, for a total of 20 schools. As its name suggests, the DIP focused primarily on reducing dropout rates, but it was also expected to improve student learning.

To determine whether or not the pilot interventions should be replicated, the government randomly assigned them to schools in select low-income communities and collected both pre- and post-intervention data over two school years. The government also collected data on schools that were not part of the program to provide a benchmark for assessing the impact of the interventions. The resulting data set is rare in a developing country.[2] Thus while one goal of this paper is to shed light on elementary education policy in the Philippines, a broader aim is to demonstrate that project evaluation in developing countries is both feasible and worthwhile.

I. THE DIP INTERVENTIONS

The DIP consisted of four experimental interventions: school feeding; multi-level learning materials, which are pedagogical materials for teachers; school feeding combined with parent-teacher partnerships; and multi-level learning materials combined with parent-teacher partnerships.[3] On a per-student basis school feeding is very expensive, the use of multi-level learning materials is considerably cheaper, and parent-teacher partnerships entail minimal additional costs because they involve mostly an adjustment in the way parents interact with teachers. The substantial differences in costs of the three interventions make it especially important to compare them in terms of both benefits and costs.

2. For recent examples of the evaluation of social sector programs based on randomized control designs, see Newman, Rawlings, and Gertler (1994). See Glewwe, Kremer, and Moulin (1997) for a recent application to education.

3. Parent-teacher partnerships envision a more active role for parents than they are commonly assigned, especially in developing countries. See Epstein (1991) for a discussion of how teachers' interactions with parents can improve student achievement.

Under the school feeding intervention all pupils in beneficiary schools received a free school meal while classes were in session. Because of substitution effects, this intervention may not increase pupils' food intake, as Jacoby (1997a, 1997b), for example, suggests. This problem is inherent to all feeding programs. Since we have no way to quantify the amount of substitution between food provided at home and food provided under the DIP, any change associated with the intervention must necessarily be interpreted as its net rather than its gross impact.

Under the multi-level learning materials intervention, all teachers in the beneficiary schools received pedagogical materials designed to help them pace their teaching according to the differing abilities of their students. Prior to implementation of the DIP, teachers attended a week-long training course on the use of the materials. Parent-teacher partnerships comprised a series of regular (usually monthly) group meetings throughout the school year between school staff and parents. The authorities chose to implement parent-teacher partnerships in combination with one of the other two interventions (rather than on its own) because the other interventions provided a way to attract parents to the meetings and provided the substantive focus for meetings.

The DIP project team, which was part of the Bureau of Elementary Education, followed a three-stage procedure in selecting schools for the interventions. They first identified five regions of the country and, within each region, two districts that met the official definition of a low-income municipality (the municipality had to meet at least three of five poverty criteria relating to education, health, housing, unemployment, and household consumption). The sample schools were located in 10 provinces: Mindoro Oriental and Palawan in Southern Luzon Region, Camarines Sur and Sorsogon in Bicol Region, Ilioilo and Negros Occidental in Western Visayas Region, Northern Samar and Western Samar in Eastern Visayas Region, and North Cotabato and Maguindanao in Central Mindanao Region. In one district the treatment choices were packaged as no intervention, multi-level learning materials, or multi-level learning materials combined with parent-teacher partnerships, while in the other district they were packaged as no intervention, school feeding, or school feeding combined with parent-teacher partnerships. The decision of which of the two intervention packages to assign to each site in the region was made by the toss of a coin.

Next, in each district the project team selected three schools that met the following criteria: each school offered all grades of instruction in the elementary cycle, with one class of pupils per grade; had a high dropout rate, based on administrative records; was not located in an area with security risks; and did not offer any school feeding services. Each school was typically the only school in its locality. Finally, by random drawing, the three schools in each district were assigned to the control group or to one of the two intervention options. The process generated a sample of 20 intervention schools and 10 control schools. One school from the control group was eventually dropped because of logistical difficulties in collecting data.

The use of random assignment yields evaluation results that are both convincing to researchers and easy for policymakers to understand (Burtless 1995). Heckman and Smith (1995) point out that selection bias may remain a problem in randomized trials because people in treatment groups may opt out of the treatment and those in the control group may compensate for their exclusion from the experiment. In the DIP evaluation, schools in the treatment group could not select into or out of the assigned interventions, and schools in the control group could not substitute other types of educational interventions to compensate for not being in the treatment group. However, because resources were scarce, the project team found it necessary to strike a balance between the priority of addressing pressing needs in poor schools and the advantage for program evaluation of having complete randomization in the placement of interventions. In the end the program team decided to target the interventions to needy districts and schools. Although randomized selection was not used to identify the two districts in each region nor the schools in each district, it was the basis for assigning the two intervention packages to the sites in each region and for assigning treatment or control status to schools.

II. Data Collection

The interventions were implemented in the 1991–92 school year, but data collection began in 1990–91 to generate baseline information (table 1). Data were collected from all pupils in all grades in each sample school. Because the schools had only one section per grade, being poor rural schools, the resulting data set has information on the full population of students.[4]

The data include the characteristics of the schools and the classroom environment (including teacher characteristics), as well as information about the pupils: family and personal background; scores on grade-specific tests in mathematics, English, and Filipino, with one set of tests administered at the start of the school year and a second set administered at the end; and transition to the next school year. The survey also attempted to record students' daily attendance throughout the school year, but the data proved unreliable because of poor record-keeping. The data on transition to the next grade comprised two kinds of information: the school management's year-end decision to promote the pupil to the next grade or retain him or her in the same grade, and whether or not the pupil actually returned the next school year.

The data are two-year records of pupils' transition through school for those who remained in the sample for both years of the study—that is, those who entered first through fifth grade in year one and who did not drop out. (There were very few transfers to other schools because most of the schools in the project sites were the only ones in their locality). The transition record is truncated for

4. We account for this feature of the sampling procedure in the multivariate regression analyses below by allowing for a school- or teacher-level structure in the variance-covariance error term.

Table 1. *Sample Composition in the Philippine Dropout Intervention Program, 1990–92*

Intervention	*Number of schools*	*Number of pupils*[a]	
		1990–91	*1991–92*
Full sample	29	4,267	3,953
Program intervention			
No intervention	9	1,356	1,279
School feeding			
Alone	5	751	695
With parent-teacher partnership	5	858	792
Multi-level learning materials			
Alone	5	673	634
With parent-teacher partnership	5	629	553

a. Includes only pupils in grades one to five with data on personal and family background who advanced to the next school year.

Source: Survey data from the 1990–92 Dropout Intervention Program.

sixth graders in year one and first graders in year two. Sixth graders had left primary school by the second year, and their schooling career was not tracked; as a result, their data are for year one only. First graders in year two have no data for year one, since they were not yet in primary school.

As with any complicated effort to collect data that involves many actors, the data that eventually became available had some shortcomings. Unlike the dropout and transition data, the data on student performance were collected for one year only. The original intention was to gather longitudinal data, but unforeseen coding problems prevented that.[5] Thus only the achievement data for year one—comprising scores at the start and at the end of the year—could be linked to the data on student background. Fortuitously, in year two the same achievement tests were administered to the new cohort of entering first graders. Adding these pupils to the first graders from year one produced a data set containing the information needed to evaluate the impact of the interventions on student learning among first graders. Since schooling outcomes in first grade turn out to be especially relevant to elementary education policy in the Philippines—that is, the dropout problem is concentrated in the first grade—the lack of suitable data for the other grades proved to be a less serious flaw than appeared at first sight.

To confirm that the DIP interventions were in fact randomly assigned across schools, we compared dropout rates and student learning as well as students' socioeconomic background in treatment schools and control schools prior to the implementation of DIP. The results suggest that in terms of the outcome variables—dropout rates and year-end test scores—the treatment schools are not significantly different from the control schools (table 2). The random assignment

5. The identification codes on the achievement files from the second year lacked sufficient detail to permit secure matches to the data from the first year. The files on attendance and transition status were collected using a separate procedure and did not suffer from this flaw.

Table 2. *Pupils in Control and Treatment Schools Prior to Implementing the Dropout Intervention Program*

Variable	Control schools	Schools that received the school feeding intervention		Schools that received the multi-level learning materials intervention	
		Alone	With parent-teacher partnership	Alone	With parent-teacher partnership
Outcome variables					
Mean dropout rate (percent)	9.56	8.58	7.02**	9.29	10.01
Mean z-score on year-end test[a]	0.02	0.01	0.07	–0.10	–0.10
Student characteristics					
Percent repeating current grade	0.21	0.28*	0.23	0.24	0.27
Percent attended preschool	19.8	14.6*	10.6**	21.1	13.5**
Percent whose father is a farmer	47.2	53.6*	47.9	36.2**	44.9
Percent from non-Tagalog-speaking homes	48.9	52.4*	57.1**	33.1**	32.0**
Mean years of mother's schooling	6.0	5.8*	5.5**	6.3**	6.1
Mean number of brothers and sisters	5.2	5.1	4.7**	5.4	4.7*
Mean z-score on entering test[a]	0.06	–0.14*	0.19**	0.15	–0.41*

* Deviation from the control group is statistically significant at the 5 percent level.
** Deviation from the control group is statistically significant at the 1 percent level.
a. Test scores are for first graders only. They are expressed in units of standard deviation from the sample mean.
Source: Authors' calculations.

process thus appears to have been valid, implying that a simple analysis of the differences between the mean impacts on the control and treatment groups would capture average treatment effects.

The schools are less similar, however, with regard to pupil characteristics. In particular, children in schools that received the school feeding program appear to be systematically less well off than children in the control schools. The presence of such differences is not surprising, given the relatively small number of schools in the sample. Below we use multivariate methods to control for these differences in evaluating the impact of the DIP interventions.

III. The Impact of the DIP Interventions

Dropout rates decline in all sample schools between the pre- and post-treatment years (table 3). However, the decline is statistically significant only in the schools that received multi-level learning materials, with or without parent-teacher

partnerships. To isolate the pure intervention effect, we compute the difference in the change in the dropout rate over time between each treatment group and the control group and then perform *t*-tests on the resulting difference-in-difference estimates. In the schools with a feeding program, for example, dropout rates decline 2.9 percentage points between year one and year two compared with a decline of 1.2 percentage points in the control group. The *t*-test on this difference-in-difference estimate (1.7 percentage points) suggests that it is not statistically significant. In contrast, in both treatment schools with multi-level learning materials (with or without parent-teacher programs) the decline in the dropout rate is statistically significant at the 10 percent level or better.

With regard to student learning in first grade, the change between the pre- and post-treatment years is particularly striking for schools that received multi-level learning materials and initiated parent-teacher partnerships, but the change is not statistically significant. Likewise, the difference-in-difference estimate, while positive and relatively large, is also not statistically significant. For the other treatment groups the change between the baseline and treatment years is more modest, and none of the estimates of the program's impacts is statistically significant. The lack of statistical significance is not surprising, however, given that we have mean test scores on only five classes of pupils for each intervention.

Table 3. *Impact of the Dropout Intervention Program on Schooling Outcomes between 1990–91 and 1991–92*

		School feeding		*Multi-level learning materials*	
Variable	*Control schools*	*Alone*	*With parent-teacher partnership*	*Alone*	*With parent-teacher partnership*
Dropout rates					
Percentage change	−1.2	−2.9	−2.8	−4.8	−6.4
P-value	0.328	0.104	0.110	0.004***	0.005***
Difference-in-difference estimate[a]	n.a.	−1.7	−1.6	−3.6	−5.2
P-value	n.a.	0.440	0.465	0.080*	0.028**
Student achievement[b]					
Change in *z*-score	0.11	−0.01	0.04	−0.17	0.47
P-value	0.787	0.989	0.910	0.809	0.240
Difference-in-difference estimate[a]	n.a.	−0.12	−0.07	−0.28	0.36
P-value	n.a.	0.839	0.902	0.705	0.500

n.a. Not applicable.
* Statistically significant at the 10 percent level.
** Statistically significant at the 5 percent level.
*** Statistically significant at the 1 percent level.
a. The difference-in-difference estimate refers to the difference between the treatment and control groups in the change in dropout rates or test scores.
b. Data are for first graders only. *z*-scores are test scores expressed in units of standard deviation from the sample mean.
Source: Authors' calculations.

These results suggest that the impact of the DIP interventions is ambiguous, positively affecting dropout behavior, but not influencing student learning. Given the systematic differences in student characteristics between the treatment and control groups, as well as the small samples involved, such a pattern is expected. Including more test sites for each of the DIP interventions would have brought more clarity, but it also would have required many more resources than the government was willing or able to allocate to the exercise. Indeed, in most developing countries large-scale experiments are rarely affordable as routine procedures for policy analysis. We can nonetheless exploit the data generated from the DIP's experimental design by conducting a multivariate analysis to control for the systematic differences between the control and intervention groups. For this analysis we use pupils rather than schools as the unit of observation.

Multivariate Analysis

As above, we focus on dropout behavior and student learning as the relevant outcomes for evaluating the DIP interventions. Following the literature (for example, Hanushek and Lavy 1994), we postulate that the probability that child i drops out of school s at time t (DP_{ist}) depends on the child's personal characteristics (PC_i) and family background (FB_i), the learning environment (LE_{st}), and the characteristics of the community (CC_i) in which he or she lives. Each DIP intervention j ($INTER_j$) can affect both dropout behavior and student learning. The school feeding intervention, for example, lowers the cost of schooling, thus boosting the incentives for parents to keep children in school. At the same time, it can improve children's general health and attentiveness, thereby stimulating academic progress and lowering the chances of dropping out. The provision of multi-level learning materials aims to improve the effectiveness of the pedagogical process. Thus it could potentially heighten children's interest in school as well as enhance their learning. Both would reduce the probability of dropping out. Finally, parent-teacher partnerships expand parents' involvement in the schooling of their children, thereby minimizing the influence of adverse social and academic factors on dropout behavior.

More formally,

$$(1) \qquad DP_{ist} = \beta_0 + \beta_1 PC_i + \beta_2 FB_i + \beta_3 LE_{st} + \beta_4 CC_i + \beta_5 INTER_{jt} + \varepsilon$$

In measuring the impact of the DIP on student learning, we again follow the literature (for example, Harbison and Hanushek 1992) in postulating that child i's academic performance in school s at time t, AP_{ist}, is a function of his or her initial achievement, AP_{ist-1}, personal characteristics (PC_i), and family background (FB_i), as well as the learning environment (LE_{st}), and community characteristics (CC_i):

$$(2) \quad AP_{ist} = \delta_0 + \delta_1 AP_{ist-1}, + \delta_2 PC_i + \delta_3 FB_i + \delta_4 LE_{st} + \delta_5 CC_i + \delta_6 INTER_{jt} + \varepsilon$$

Two important econometric issues—discussed fully in Angrist and Krueger (1999) and Vella (1998)—arise in the estimation of equation 2. The first is the

presence of a lagged dependent variable, which, while providing an important control factor, may be correlated with the error term. This problem can be addressed by choosing instrumental variables that are correlated with the lagged variable but not with the error term. We choose the lagged values of scores on other tests as instruments. Of course, the reduction in bias comes at the expense of a loss of efficiency, and the reliability of this approach depends on the validity of the instruments. In our case the *r*-squared values of the correlation between the instruments and the lagged values range from 0.43 to 0.50, suggesting some loss of efficiency and a consequent downward bias in the *t*-statistics.

The second econometric issue is that of selection bias associated with dropping out. Since the weakest students are those who are most likely to drop out, the analysis of student learning is performed on a censored sample. Although the dominant method of correcting for this problem is to apply a Heckman two-step correction by constructing an index based on the probability of censoring, there is some dissatisfaction with this approach, as Hamermesh (1999) describes.[6] The index is based on the assumption of a normal distribution, and, as Hamermesh notes, the results are extremely sensitive to distributional assumptions. Further concerns are often raised about the choice and adequacy of identifying variables in the first step, although in our case we are fortunate in having detailed information on variables in equation 1 that affect the cost of education but are not directly associated with a child's academic performance. In particular, we have data on the distance to school, whether or not the father is a farmer (since an important opportunity cost of schooling in poor rural communities is children's contribution to farm work), and whether or not the student is the oldest child in the family.

A number of alternative approaches have been proposed to deal with selection bias—nonparametric and semiparametric methods, as well as nonindex-oriented models—although no consensus has yet emerged as to which is preferred. Consequently, we estimate and report the results from three separate procedures.[7] As a basis for comparison, we first report the results of a simple regression of year-end test scores against the intervention variables and control factors, with no correction for selection bias. Then we use a nonindex instrumental variable approach, following Krueger (1997), in which we simply assign to students who have dropped out their academic ranking based on their initial test score. Finally, we follow the conventional Heckman approach, which includes the Mills ratio as an additional regressor in equation 2. It should be noted that in the Heckman approach the standard errors are biased because it is impossible to estimate them correctly when simultaneously applying that procedure and using instrumental variables to correct for the problem of lagged dependent variables.

6. Noteworthy, however, is Vella's (1998) finding that the Heckman approach performs quite well compared with other approaches.

7. In all three approaches we correct for the problem of having a lagged dependent variable by using the instrumental variables procedure described above.

To estimate equation 1 we use data for pupils from the first to fifth grades in order to increase the sample size for analyzing what in statistical terms is still a relatively rare event. We could not include pupils from the sixth grade because their dropout record is incomplete for reasons explained earlier. We represent a pupil's characteristics and family background as a vector of commonly used variables, such as the child's sex and mother's education.[8]

Regression Results

The full dropout regression model, with controls for the complete range of background factors, achieves a reasonable overall goodness-of-fit, with a Hosmer-Lemeshow chi-squared statistic of 13.28 (column 2 of table 4). We also estimate a simplified model using only the interventions as regressors in order to see whether the interventions were correlated with the background variables (column 3 of table 4).[9] Not surprisingly, the simplified regression model as a whole has no explanatory power; it is nonetheless noteworthy that all of the coefficient estimates are comparable to the corresponding estimates in the full model. Moreover, in both regressions only the intervention involving the use of multi-level learning materials has a measurable effect on dropout behavior. The positive impact of this intervention is consistent with the results based on sample means (see table 3). Those results also suggest that interventions combining the use of multi-level learning materials with parent-teacher partnerships have a positive impact.

We then estimate equation 2 for the three different subjects—mathematics, Filipino, and English. Students were given tests in these subjects at both the start and end of the school year (table 4).[10] The first feature of the results is that no intervention consistently improves student learning across all three subjects, a finding that jibes with the data in table 3, which are based on average performance across the three subjects. For both Filipino and English the coefficients on the intervention involving multi-level learning materials combined with parent-teacher partnerships are statistically significant in all three regressions. The coefficients are comparable in magnitude, especially in the regressions for English.

For English the coefficients on the school feeding intervention, whether alone or combined with the parent-teacher partnerships, are statistically significant only in the two regressions that control for selection bias (the second and third columns in each subject block). For mathematics the coefficient on school feeding combined with parent-teacher partnerships is also statistically significant in both regressions adjusted for selection bias. Overall, the findings suggest that the DIP interventions are better at helping students learn languages than mathematics. Further, the interventions involving the use of multi-level learning materials, as

8. Father's education is almost perfectly collinear with mother's; we use the latter because the mother is more likely to provide help with homework.

9. We thank an anonymous referee for suggesting this specification.

10. In addition to the variables mentioned above, we also include teacher and school fixed effects. For each subject we use incoming test scores on the other two subjects to instrument the incoming test score.

currently designed and implemented, appear to produce more consistent results than do the other components.

IV. Policy Implications

To assess the policy implications of the DIP, we need to consider both its cost and impact. We can assess these qualitatively with available cost data and our interpretation of the regression results discussed above (table 5). The underlying cost data are drawn from implementation records kept by the Bureau of Elementary Education. The school feeding program costs an average of P946 (pesos) per beneficiary, with a range between P621 and P1,054. The multi-level learning materials program involved only pedagogical materials, which cost an average of P90 per child, and the parent-teacher partnerships involved monthly meetings that cost an average of about P33 per child (in direct costs). To be perfectly comparable, these costs should be adjusted for the opportunity cost of time—that of teachers supervising the school feeding program, that of parents and teachers in parent-teacher partnerships, and that of people who train teachers to use the multi-level learning materials. The cost of the multi-level learning materials program should also be adjusted to reflect the fact that the pedagogical resources it provides have a typical lifetime of more than one year. Without making the adjustments explicit, however, the existing cost data already point to an obvious ranking of the interventions, with school feeding at the high end, followed by multi-level learning materials and parent-teacher partnerships at the low end.

The impact of the interventions on dropout behavior and student learning range from nonexistent (0) to promising (++++). The ranking reflects our interpretation of the results from the difference-in-difference estimates and the regression analysis. If none of the analyses shows an appreciable impact, we categorize the intervention as having zero impact; if the results are consistently positive across all or most of the estimation methods and the magnitude of the impact is relatively large, we label the intervention promising and assign it three or four pluses to reflect the degree of consistency; and if the results are sporadically positive, we categorize the intervention as having a weak impact, assigning it only one or two plus signs.

Given these cost and benefit criteria, the combination of multi-level learning materials and parent-teacher partnerships appears to be the most cost-effective. In contrast, the school feeding intervention, at least in the form implemented in DIP, seems to be a weak candidate for replication. This does not imply that a more targeted program—directed, for example, at only the most malnourished and underprivileged children—would not be cost-effective. That possibility cannot be confirmed, however, with the data available here. Note, though, that the impact on student learning of multi-level learning materials in combination with parent-teacher partnerships is probably limited to instruction in languages.

Table 4. *Regression Results of the Impact of the Dropout Intervention Program*

			Year-end test scores (first graders only)								
			Math			*Filipino*			*English*		
				Correction for selection bias			*Correction for selection bias*			*Correction for selection bias*	
Independent variable	*Probability of dropping out (first to fifth graders)* (1)	(2)	*No correction for selection bias*	*Using nonindex instrumental variable approach*	*Using Heckman's approach*	*No correction for selection bias*	*Using nonindex instrumental variable approach*	*Using Heckman's approach*	*No correction for selection bias*	*Using nonindex instrumental variable approach*	*Using Heckman's approach*
Intervention variables[a]											
School feeding	−0.254	−0.255	0.241	0.248	0.121	0.317	0.160	−0.031	0.317	0.323	0.009
	(0.56)	(1.26)	(0.77)	(2.72)**	(1.36)	(1.80)	(1.82)	(0.32)	(1.80)	(3.63)**	(3.73)**
Multi-level materials	−0.428	−0.458	−0.092	−0.045	−0.008	0.647	0.234	0.178	0.647	0.548	0.543
	(1.71)*	(1.99)*	(0.18)	(0.38)	(0.07)	(1.66)	(2.05)*	(1.42)	(1.66)	(4.71)**	(0.66)
School feeding with parent-teacher partnerships	−0.311	−0.319	0.370	0.347	0.277	0.458	0.114	0.058	0.458	0.442	0.544
	(1.40)	(1.63)	(0.84)	(3.74)**	(3.08)**	(1.63)	(1.28)	(3.31)**	(1.63)	(4.89)**	(1.66)*
Multi-level materials with parent-teacher partnerships	−0.410	−0.367	0.217	0.210	0.081	0.870	0.225	0.309	0.870	0.754	1.048
	(1.15)	(1.56)	(1.50)	(1.83)	(0.76)	(3.12)**	(2.02)*	(2.65)**	(3.12)**	(6.64)**	(8.83)**
Initial test scores (instrumented)[b]			0.510	0.607	0.520	0.373	0.618	0.473	0.373	0.485	0.341
			(8.69)**	(18.07)**	(15.16)**	(5.11)**	(18.51)**	(13.98)**	(5.11)**	(13.83)**	(9.99)**
Pupil is a girl	−0.172		0.153	0.116	0.171	0.290	0.224	0.246	0.290	0.240	0.181
	(1.61)		(2.64)*	(2.95)**	(4.43)**	(4.57)**	(5.89)**	(4.46)**	(4.57)**	(6.24)**	(2.08)*
Attended preschool	−0.211		−0.151	−0.173	−0.096	−0.014	−0.045	0.468	−0.014	−0.039	0.500
	(1.08)		(2.12)*	(2.96)**	(1.61)	(0.25)	(0.79)	(3.57)**	(0.25)	(0.69)	(1.18)
Mother's years of education	−0.072		0.013	0.016	0.007	0.001	0.010	0.004	0.001	0.000	−0.006
	(3.22)**		(1.16)	(2.11)*	(0.87)	(0.15)	(1.32)	(0.44)	(0.15)	(0.03)	(2.22)*
Father is a farmer	−0.070		0.026	0.024	−0.035	−0.103	−0.002	0.143	−0.103	−0.122	−0.081
	(0.52)		(0.26)	(0.47)	(0.74)	(0.86)	(0.04)	(1.28)	(0.86)	(2.44)*	(1.52)
Non-Tagalog speaker	−0.140		−0.031	0.000	0.028	−0.005	0.183	0.014	−0.005	0.009	0.034

Eldest child	0.092										
	(0.70)										
Repeated previous grade	0.327										
	(2.68)**										
Child has active personality[c]	−0.133										
	(0.62)										
Family income (pesos per year)	−0.109										
	(1.24)										
Distance to school (kilometers)	0.450										
	(0.88)										
Distance squared	−0.125										
	(0.97)										
Hosmer-Lemeshow χ^2	13.28	0.00									
(p-value)	(10.25)	(1.00)									
Inverse Mills ratio					−0.087			−0.351			−0.471
(standard errors)					(0.178)			(0.122)			(0.099)
Number of observations	8,229	8,229	1,676	1,676	1,676	1,676	1,676	1,676	1,676	1,676	1,676
R-squared			0.41	0.41		0.43	0.45		0.42	0.44	

* Statistically significant at the 10 percent level.

** Statistically significant at the 5 percent level.

*** Statistically significant at the 1 percent level.

Note: The t-statistics (in parentheses) are consistent with standard errors adjusted for group-specific heteroskedasticity using the Huber-White correction procedure. For the dropout regression the standard errors are adjusted for clustering on schools. For the test score regressions the standard errors are adjusted for clustering on teachers. The dropout regressions include dummy variables for grades.

a. The intervention variables are defined as dummy variables that take on the value of 1 when the school attended by the child is a recipient of the indicated intervention and zero otherwise.

b. Tests scores at start of the school year, instrumented by scores on the other two subjects.

c. As assessed by pupil's teachers.

Source: Authors' calculations.

Table 5. *Policy Implications of the Dropout Intervention Program Evaluation*

Intervention	*Costliness of intervention*	*Impact of intervention*[a]	
		On dropout behavior	*On test scores*
School feeding	High	0	+ +
Multi-level learning materials	Low	+ + + +	+
Parent-teacher partnership with school feeding	High	0	+ +
Parent-teacher partnership with multi-level learning materials	Low	+ +	+ + + +

a. A rating of 0 indicates no impact, one or two pluses indicates a weak impact, and three or four pluses indicates a strong impact.

Source: Authors' calculations.

V. Conclusions

The DIP represents an important effort by the Philippine government to experiment with new ways to address problems in elementary education. To evaluate the program, the government collected pre- and post-treatment data on pupils in test schools and in control schools. We used these data here to assess how the DIP interventions affect dropout behavior and learning. Data coding problems limited the analysis of student achievement to first graders only.

Dropout rates and student achievement in the control and treatment schools are comparable in the baseline year, suggesting that the random assignment process worked as expected. However, the schools differed in the background characteristics of pupils, so that a simple comparison of mean differences between the control and treatment schools before and after implementing the DIP was not sufficient to establish the true impact of the interventions. The fact that the sample included only five classes of pupils per intervention also hampered the analysis. Both deficiencies prompted us to use multivariate analysis to complement the comparison of means in the control and treatment schools.

The evaluation period is admittedly too short to reach firm conclusions, but the preliminary findings reported in the paper offer a good basis for assessing the economic justification for replicating the DIP interventions. Taken as a whole, they imply that, of the four interventions implemented, the argument for replication is strongest for multi-level learning materials combined with parent-teacher partnerships, and weakest for the school feeding program, at least as it was implemented in the DIP.

It is important to note that if improved student learning is an objective, the combination of multi-level learning materials and parent-teacher partnerships is only one of many potential interventions (such as offering preschool education, expanding teacher training, improving classroom conditions, and supplying more student textbooks and workbooks). Thus although our evaluation offers some support for replicating one of the DIP experiments, it by no means implies that we have finished our search for ways to address dropout and student learning prob-

lems in the Philippines. Further, given that the experiment was effective mainly in promoting student performance in Filipino and English, other measures clearly need to be considered if improved performance in mathematics is also desired.

The search for cost-effective strategies to improve schooling outcomes is an issue in all countries. Often, however, such work is hampered by the absence of a routine system for assessing alternative investment options. The fact that the DIP was implemented and evaluated augurs well for the future, the problems encountered notwithstanding. It shows that the institutional capacity to evaluate social experiments properly exists or can be nurtured within ministries of education or related agencies. The task of building research capacity, particularly when the institutions involved are outside of academia, is undeniably difficult. The benefits are probably worth the effort, however, because the scope for mistakes in the choice of investment decisions is wide in the absence of quantitative information about education processes. And these mistakes are costly, not only in financial terms, but also in terms of hindering children's education.

REFERENCES

The word "processed" describes informally reproduced works that may not be commonly available through library systems.

Angrist, Joshua, and Alan Krueger. 1999. "Empirical Strategies in Labor Economics." In Orley Ashenfelter and David Card, eds., *Handbook of Labor Economics*. Amsterdam: North-Holland Press.

Burtless, Gary. 1995. "The Case for Randomized Field Trials in Economic and Policy Research." *Journal of Economic Perspectives* 9(2):63–84.

Chuard, Dominique, and Alain Mingat. 1996. "Analysis of Dropout and Student Learning in Primary Education in South Asia." Asian Development Bank, Manila. Processed.

Epstein, Joyce L. 1991. "Effects on Student Achievement of Teacher Practices of Parent Involvement." In Steven Silvern, ed., *Advances in Reading/Language Research*. Vol. 5: *Literacy through Family, Community, and School Interaction*. Greenwich, Conn.: JAI Press.

Glewwe, Paul, Michael Kremer, and Sylvie Moulin. 1997. "Textbooks and Test Scores: Evidence from a Prospective Evaluation in Kenya." Development Economics Research Group, World Bank, Washington, D.C. Processed.

Gomes-Neto, Joao Batista, and Eric A. Hanushek. 1994. "The Causes and Consequences of Grade Repetition: Evidence from Brazil." *Economic Development and Cultural Change* 43(1):117–48.

Hamermesh, Daniel S. 1999. "The Art of Labormetrics." In Orley Ashenfelter and David Card, eds., *Handbook of Labor Economics*. Amsterdam: North-Holland Press.

Hanushek, Eric A., and Victor Lavy. 1994. "Schooling Quality, Achievement Bias, and Dropout Behavior in Egypt." Living Standards Measurement Study Working Paper 107. Development Economics Research Group, World Bank, Washington, D.C. Processed.

Harbison, Ralph W., and Eric A. Hanushek. 1992. *Educational Performance of the Poor: Lessons from Rural Northeast Brazil*. New York: Oxford University Press.

Heckman, James J., and Jeffrey A. Smith. 1995. "Assessing the Case for Social Experiments." *Journal of Economic Perspectives* 9(2):85–110.

Jacoby, Hanan. 1997a. "Is There an Intrahousehold 'Flypaper Effect'? Evidence from a School Feeding Program." Jamaica School Feeding Program. Processed.

———. 1997b. "Self-Selection and the Redistributive Impact of In-Kind Transfers." *Journal of Human Resources* 32(2):233–49.

Krueger, Alan. 1997. "Experimental Estimates of Educational Production Functions." NBER Working Paper 6051. National Bureau of Economic Research, Cambridge, Mass. Processed.

Lockheed, E. Marlaine and Adriaan M. Verspoor. 1991. *Improving Primary Education in Developing Countries*. Washington D.C.: Oxford University Press for the World Bank.

Miguel, Marcelina M. 1993. "Variables That Affect Student Achievement in the Philippines." Bureau of Elementary Education, Department of Education, Culture, and Sports, Manila. Processed.

Newman, John, Laura Rawlings, and Paul Gertler. 1994. "Using Randomized Control Designs in Evaluating Social Sector Programs in Developing Countries." *The World Bank Research Observer* 9(2):181–201.

Tan, Jee-Peng, Julia Lane, and Paul Coustère. 1997. "Putting Inputs to Work in Elementary Education: What Can Be Done in the Philippines?" *Economic Development and Cultural Change* 45(4):857–79.

Vella, Francis. 1998. "Estimating Models with Sample Selection Bias: A Survey." *Journal of Human Resources* 33(1):127–69.