IDB Education

BRIEFLY NOTED

Nº 4 FEBRUARY 2010

Ana Santiago
Marcelo Cabrol
Mariana Alfonso

ALIGNING LEARNING INCENTIVES (ALI)

Summary: Mexican students score low in international learning assessments, particularly in math. To help remedy this situation, the Inter-American Development Bank is supporting a pilot study to identify to what extent performance based incentives for students, teachers and principals in upper-secondary schools can improve students' mathematical skills, as measured by their scores in curriculum-based mathematics tests.

Three alternative incentive designs are being examined during the 2008/09 through 2010/11 academic years. In the first design, rewards are provided only to students based on their performance and improvements in their performance on the mathematics examinations. In the second design, rewards are provided only to teachers based on their students' performance on the same examinations. In the third design, rewards are provided to students, teachers, and school administrators based on the students' performance on the same examinations. In this latter design, extra rewards are provided to students if their classmates perform well and to teachers if the students in other mathematics classes in the school do well. In this monograph we describe the pilot initiative and evaluation design, while also summarizing some of the baseline data and background information. Preliminary results suggest that the third design in which incentives are aligned among students, teachers and principals results in significant increases in mathematics performance.

CONTENTS

Inter-American Development Bank
1300 New York Avenue, NW
Washington DC 20577 USA

Low Rankings in International Assessments

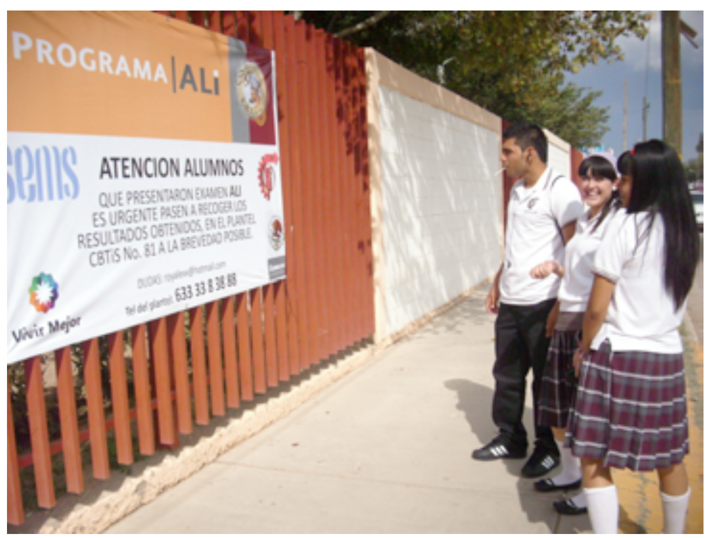
Controlling for per capita GDP, students from Latin America and the Caribbean generally, and from Mexico in particular, consistently perform below students from the industrialized countries that make up the Organisation for Economic Co-operation and Development (OECD) in international assessments of educational achievement. In 2006, the Latin American countries participating in the OECD Programme for International Student Assessment (PISA), which focuses on secondary education, were all among the bottom performers. The situation is particularly challenging in the area of mathematics, a basic competency for competing in a globalized economy.

While Mexico has made significant progress in widening access to education and in expanding coverage of the school-age population, and some progress in the mathematics PISA scores, the 2006 PISA results are still disappointing. Mexico ranked last among the OECD countries and 48th among all participating countries in mathematics out of 57 participating countries.

To address this pervasive problem, governments are seeking new models to help improve student learning. One of such models is the introduction of incentive schemes that seek to change behavior and align all education actors towards one common goal: improve student achievement.

ALI Design

The main objective of the Aligning Learning Incentives (ALI) program is to evaluate alternative monetary incentive designs aimed at improving student achievement in Mathematics at the upper secondary level (10th, 11th and 12th grades) in Mexico. An important rationale for utilizing monetary incentives is that Ministries of Education or similar centralized decision makers are not likely to know the best means of improving education in the heterogenous and multi-dimension conditions that students, teachers and principals face. But by providing incentives that are tied directly to curriculum-based examinations, students, teachers and principals can be induced to choose the best means to improve performance given their specific circumstances.



The ALI program, which extends over a three-year period, covers three sets of 20 schools, each set subject to a different incentive design, and a fourth set of 28 schools with no incentives. Broadly, the incentive designs differ with respect to whether students, teachers and administrators – or all three groups, are rewarded. Specifically, the four groups are:

1) Treatment group 1 (T1): Performance rewards to students who achieve well-defined achievement targets on standardized grade-specific curriculum-based Mathematics examinations developed for this project.
2) Treatment group 2 (T2): Performance rewards to Mathematics teachers based on the performance of the students in their classes on standardized grade-specific curriculum-based Mathematics examinations developed for this project.
3) Treatment group 3 (T3): Performance rewards to students, all teachers (Mathematics and non-Mathematics) and school administrators based on the performance of students in the school on standardized grade-specific curriculum-based Mathematics examinations developed for this project. To encourage synergies among students, student rewards depend both on their own and on their classmates' performance. To encourage synergies among teachers, teacher rewards depend both on the performance of their own students and of all other students in the school.
4) Control group (C): No performance rewards.

The level and improvements in mathematics performance are measured by mathematics tests designed by CENEVAL (Cento Nacional de Evaluación para la Educación Media Superior, National Center of Evaluation for Upper Secondary School) for this project. The design of the instruments benefited from systematic inputs of Mexican experts on upper secondary school mathematics and pilot testing. These tests are explicitly based on the Mexican curriculum, including changes experienced in recent reforms. These tests are administered annually by the Ministry of Education with procedures designed to assure independent evaluations and minimize possible testing abuses.

Pilot Evaluation Design

ALI has been designed for an experimental evaluation. The evaluation sample consists of 88 Federal upper secondary schools. These schools are a selected subset of the 706 Federal upper secondary schools that satisfy the following criteria: (i) not in their first year of operation, (ii) only one session (in the morning) per day, (iii) technically oriented agricultural and industrial schools , (iv) no other federal upper secondary school within 10 miles, (v) between 200 and 2000 students, (vi) not in two states which were excluded due to feasibility constraints, and (vii) only one campus. As of the Fall 2008 semester, these schools included over 48,000 students and over 400 math teachers.

These 88 schools were randomized into four groups, 20 in each of the three treatment groups and 28 in the control group. Randomization was performed using a school-based block randomization design (Cox and Reid 2000), where all students and/or teachers within a school were included in the same treatment or control regime. The rationale for blocking is to improve precision by using prior knowledge on which baseline characteristics are likely to be associated with the treatment responses. For maximum benefit, units should be grouped into blocks so that all units within a block might be expected to give similar responses in the absence of treatment differences. The schools were first grouped into 9 blocks, where the block definitions were based on the school size and the average graduation rate for the most recent year within each school. Within each block, schools were allocated at random to treatment regimes. The block definitions (cut-offs on school size and graduation rates) were chosen to have roughly similar numbers of schools within each block, although there is still some variation in total number of schools across blocks.

After creating a randomized set of schools allocated to treatment/control regimes, the comparability across groups in terms of the following baseline characteristics that were not used as blocking variables was examined: percentage of teachers with university degrees, percentage of new directors (principals), regional distribution, school type (DGETI or DGETA), percentage of Oportunidades recipients within schools, distances to nearest Federal upper secondary schools and mean class sizes. Following the recommended procedure of Cox and Reid (2000), several (six) randomizations were undertaken and the randomization was selected in which the four groups (T1, T2, T3 and C) were most comparable in terms of mean baseline observable characteristics. This procedure is preferable to the alternative of correcting imbalance in covariates via regression adjustment after randomization because that alternative leads to a loss in degrees of freedom and inflates the variance of the estimated treatment effect. Table 1 gives the means for characteristics on which blocking was undertaken and the characteristics on which balance was checked. With the final randomization each of the three treatments did not differ significantly at the standard 5% level from the controls for any of these characteristics.

Baseline Findings

The ALI program was initiated early in the 2008 academic year. Prior to program initiation, and before notifying the schools about the program, extensive survey information was collected from all students and all Mathematics teachers in the 88 schools in the evaluation sample. This survey covered various baseline characteristics and previous study and teaching practices. The reasons for collecting this information are: 1) to assess whether there is balance among the four groups (T1, T2, T3, C), 2) to permit greater precision in the estimation of program effects and 3) to establish baseline information on studying and teaching practices in order to learn if these practices changed as a result of the different incentive designs.

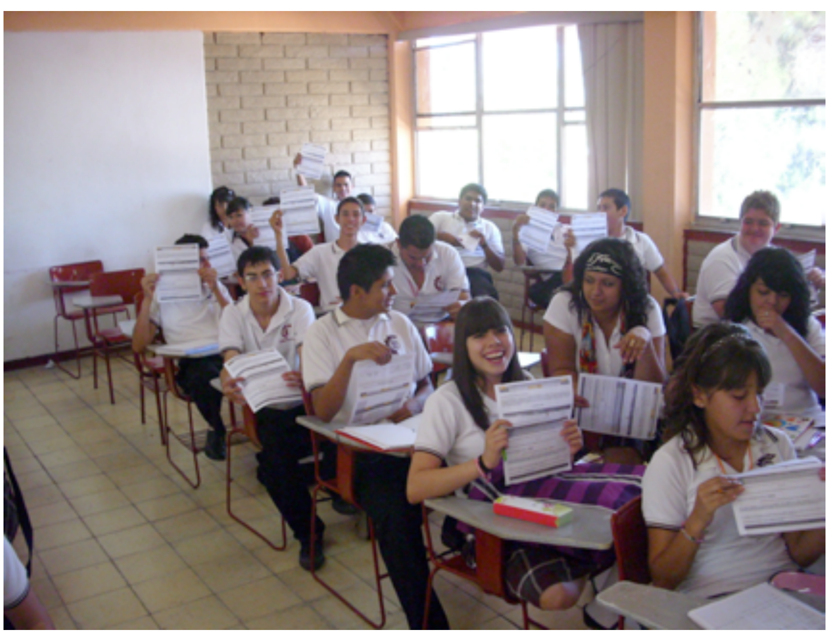
Principals for T1, T2 and T3 schools and several Mathematics teachers for each T2 and T3 schools were informed about the incentive designs in which they would be involved at meetings held at the Ministry of Education. Separate sessions were carried out for each of these three treatment groups. Students in T1 and T3 schools were then informed about the program through grade-specific and program-specific videos.

The baseline data permit description of some important characteristics of schools (Table 1), students (Table 2) and of Mathematics teachers (Table 3). For the comparison among the four groups, the important result, as noted above for Table 1, is that for all of these variables the three treatment groups do not differ significantly from the control group at standard significance levels. But in addition, of course, the means of these variables are of interest in themselves because they reveal characteristics of the population being studied.

The Federal upper secondary schools being studied on average have about 600 students each, with a mean class size of about 35-40 students. About four-fifths of the teachers have University degrees. Substantial shares of the students – about two-fifths – come from poor families that are incorporated into the Oportunidades conditional cash transfer program. Of all the students who enter the 10th grade in these schools, about three-fifths graduate from the 12th grade.
About half of the students are male. The 10th grade students in 2008-9 averaged about a quarter of a standard deviation above the national mean in the national 9th grade mathematics test (ENLACE). Though these students tend to be better than the national averages, only a relatively small percentage (about 8 percent) received satisfactory or advanced scores on these examinations, which seems consistent with the international comparisons noted above.

As seen in the table, only about a quarter of the students spent an average of four hours or more per week on Mathematics homework during the previous year, somewhat over half of the students have ever worked, and over two fifths worked in the previous school year.

The Mathematics teachers, in contrast to the students, are primarily male (over 80 percent). On average they are in their late forties, with about two decades of teaching experience, about a quarter of which is at their current school. They averaged almost four Mathematics courses in their university education. In the previous year they assigned homework about three days a week, with about a third of this homework requiring more than one hour of the students' time.



At the end of the 2008 academic year, the mathematics tests were administered to the students and further questionnaires were administered to the students and teachers in all four groups of schools. Incentive payments for T1 and T3 students, T2 and T3 teachers and T3 administrators who earned them were made via bank or credit card arrangements later in 2009.

Table 1. Comparison of Treatment and Control Schools (2007-2008)

	C	T1	T2	T3
Blocking Variables				
Mean Number of Students	582	632	609	550
Mean Graduation Rate (%)	58.3	60.4	56.2	57.9
Other Variables				
% Oportunidades	40.3	39.5	40.6	40.1
Mean Class Size	35.8	41	39	35.7
% Teachers with University Degree	82.3	79.4	81.7	84.8
Km. to Closest Fed Upper Sec School	32.9	32.8	31.4	32.4
% New Principals	25	25	30	40
% Industrial Technical Schools	46	50	55	45
% Region 1	35.7	35	50	50
Region 2	39.3	45	40	35
Region 3	17.9	10	5	10
Region 4	7.1	10	5	5

Table 2. Comparison of Treatment and Control Students (2007-2008)

	C	T1	T2	T3
% Male	49.5	48.7	48.8	48.4
9th Grade ENLACE Test Score in Mathematics (2008 10th grade class)	524.6	529	521.3	535
%. Ever worked	53.8	52.6	58.6	56.2
% Worked during previous school year	42.5	42.4	44.6	43.8
% Spent 4 hours/week or more on math homework in previous school year	25.8	26.4	27	27.2

Table 3. Comparison of Treatment and Control Teachers (2007-2008)

	C	T1	T2	T3
% Math teachers female	13	17.8	17.1	13.8
Mean age of math teacher	48.3	46.5	49.1	46.3
Mean number of math classes taken in college	3.7	3.9	3.9	3.7
Mean years of teaching experience	21.1	17.9	22.7	17.8
Mean years of teaching at current school	4.5	4.1	5	4.1
Number of days/week assigned homework				
10th grade	2.8	2.8	2.8	2.6
11th grade	2.9	2.8	3	2.6
12th grade	2.9	3	3	2.6
% time required for homework assignment more than one hour				
10th grade	33.6	27.5	29	21.5
11th grade	34.4	30.1	30.3	29.3
12th grade	36.7	34.5	30.2	26.6

Conclusion

Within an international perspective such as that provided by the PISA examinations or in comparison with the standards established by Mexican educators on the ENLACE, Mexican students perform poorly in Mathematics. In today's knowledge-based societies, understanding basic mathematical concepts is more critical than ever. Poor skills in mathematics are likely to be a serious disadvantage for any individual who wishes to function effectively in modern society. On the national level, they may constitute a major economic cost in terms of lost productivity and reduced international competitiveness.

The ALI program is an innovative program to assess, in the context of Mexican upper secondary schools, whether alternative incentive programs might have substantial impact on student mathematical performance. The provision of significant financial incentives to students, teachers and administrators in pursuit of such a goal has not been investigated before in Mexico or elsewhere. The undertaking of the three year evaluation of this program, with the alternative incentive structures, has considerable promise for learning about how powerful such incentives are and how important are synergies among the various participants in the learning processes. The baseline data are informative about some of the limitations under which Mexican upper secondary school students are attempting to improve their performance in Mathematics. The experimental design with collection of substantial baseline and follow-up data promises insights about not only whether the program works but how it works. Preliminary results from the first year indicate definite promise of the incentive design in which incentives are aligned for students, teachers and administrators (T3) as compared with alternatives of incentives limited to students (T1) or teachers (T2) or no incentives. But because these results are based only on the first year of this new program, and because the incentive effects may be cumulative over the whole upper secondary school experience, it is important to wait until the ALI results based on three years are analyzed to come to a conclusion about the size of the effects of this exciting and innovative program.

References

Behrman, J.R., S.W. Parker, P.E. Todd and K.I. Wolpin (2010): "ALIneando Incentivos para el Aprendizaje (ALI) to Improve Upper Secondary School Student Achievement in Mathematics in Mexico: Preliminary Report for IADB." Philadelphia, PA: University of Pennsylvania.

Cox, D.R. and N. Reid (2000): The Theory of the Design of Experiments,
Chapman and Hall/CRC Press.

About the Authors

Marcelo Cabrol is Chief of the Education Division of the IDB
Ana Santiago is an Evaluation Economist at the IDB
Mariana Alfonso is Education Specialist at the IDB