USAID
FROM THE AMERICAN PEOPLE

EdData II

MEASUREMENT AND RESEARCH SUPPORT TO EDUCATION STRATEGY GOAL 1

INCENTIVES TO TEACH, INCENTIVES TO READ: A PILOT OF SYMBOLIC INCENTIVES FOR TEACHERS AND STUDENTS IN JORDAN

SEPTEMBER 2016

This publication was produced for review by the United States Agency for International Development. It was prepared by RTI International.

EdData II

MEASUREMENT AND RESEARCH SUPPORT TO EDUCATION STRATEGY GOAL 1

INCENTIVES TO TEACH, INCENTIVES TO READ: A PILOT OF SYMBOLIC INCENTIVES FOR TEACHERS AND STUDENTS IN JORDAN

Education Data for Decision Making (EdData II) Project
Task Order No. AID-OAA-BC-12-00003
RTI Task Order 20, Activity 10 (Project No. 0209354.020)

Prepared for

Office of Education
Bureau for Economic Growth, Education, and Environment (E3)
United States Agency for International Development (USAID)
Arthur Muchajer, Contracting Officer
Penelope Bender, Contracting Officer's Representative

Prepared by

Lee Nordstrum and Aarnout Brombacher
RTI International
3040 East Cornwallis Road
P.O. Box 12194
Research Triangle Park, NC 27709-2194
USA

DISCLAIMER

Table of Contents

List of Figures

List of Tables

Abbreviations

CCT	conditional cash transfer
CPD	continuing professional development
E3	USAID Bureau for Economic Growth, Education, and Environment
EdData II	Education Data for Decision Making project
EGMA	Early Grade Mathematics Assessment
EGRA	Early Grade Reading Assessment
FFE	food for education
MENA	Middle East and North Africa region
RAMP	Reading and Math Program (Jordan)
RTI	RTI International (registered trademark and trade name of Research Triangle Institute)
SD	standard deviation
SE	standard error
SES	socioeconomic status
SSME	Snapshot of School Management Effectiveness
THR	take-home rations
USAID	United States Agency for International Development

1. Introduction

Motivation is key to behavior change in teaching and learning processes. Motivated teachers are thought to be more likely to be willing to experiment with different instructional approaches in the classroom. Motivated students are hypothesized to put forth extra effort during learning activities. These assumptions posit that incentives, or rewards based on evidence of behavior change, may have a role to play in enhancing the motivation of teachers and students.

This report presents findings from a short-term incentive pilot program conducted in one governorate in Jordan. During this pilot, students in treatment schools were offered a symbolic (non-monetary) incentive if they read at least 24 books at home over the 8-week implementation period. Teachers were offered a symbolic incentive if they received high scores from coaches during most (at least 50%) of their observed classroom lessons. This report presents findings from this pilot program.

2. Literature review

2.1 Teacher incentives and motivation

Motivation is essential for teachers to do what they are supposed to do well, but motivating teachers requires identifying and implementing appropriate incentives. Additionally, measuring a teacher's response to an incentive is crucial to determine whether the incentive was effective. If a teacher receiving an incentive fails to be motivated to improve his/her performance, it is logical to ask whether (1) the incentive was not appropriate or (2) the problem is unrelated to the provision of incentives (Arcia 2014).

Recently, countries in the Middle East and North Africa (MENA) expanded educational access significantly and increased the technical requirements for becoming a teacher. However, many of these same countries failed to address teacher motivation through incentives. Instead, the education systems became highly centralized, authoritarian, and bureaucratic, with little accountability to parents (World Bank 2010). Teacher incentives in the region were based on salary differentials for teachers in rural areas and by seniority, but their impact on teacher effort was limited.

Currently, analysts agree that the successful implementation of teacher incentives in MENA should include more competition in the provision of education by fostering more private education (Baum et al. 2014), increased school autonomy and school accountability to create a new social contract between teachers and parents (Demas and Arcia 2015), and better professional development paths for teachers (Brixi, Lust, and Woolcock 2015). Understanding how these suggestions relate to teacher motivation

within this restrictive environment is very important to design and implement a continuing professional development (CPD) program.

This literature review has two components. The first one is a review of global experiences with teacher incentives, illustrating what works and what does not in improving student outcomes. The second part addresses the issue of teacher motivation in relation to professional development programs.

2.1.1 Teacher incentives

Teacher incentives come in many forms, and experience in numerous countries shows that they can be likened to a basket of goods that, in addition to more money, may include recognition and prestige in the community, job stability, pension benefits, professional growth, good school facilities, and personal growth. A simple classification of incentives suggests three basic types:

1. *Professional rewards*, including prestige, intrinsic satisfaction, public recognition, professional growth, intellectual mastery, and pleasant working conditions (Bruns and Luque 2014; Mourshed, Chijioke, and Barber 2010; Vegas and Umansky 2005);
2. *Accountability pressures* (parents, peers, supervisors, and threats of demotion/dismissal), which seem to work well in decentralized education systems (Bruns, Filmer, and Patrinos 2011; Arcia et al. 2011); and
3. *Financial incentives,* such as salary, bonuses, and retirement benefits (Arcia 2014; Vegas and Umansky 2005; Bruns and Luque 2014).

Professional rewards can be effective, but they must be designed very carefully. In principle, professional rewards are given to a subset of teachers who are already motivated or have performed well. In this regard, professional rewards tend to be given *a posteriori;* that is, teachers must first improve their performance before they are recognized by the school. Intuitively, as Vegas and Umansky (2005) argued, many teachers perform well because of the intrinsic reward that they associate with teaching, their sense of mastery over their teaching subjects, and the personal satisfaction associated with imparting knowledge. However, such teachers are already motivated. The rewards they receive are simply a recognition of their motivation. Intuitively, these teachers would require little further incentive in exchange for participation in CPD.

Accountability pressures are more direct than professional rewards because they link teacher performance to pressure from parents, administrators, and peers. Accountability pressures are associated with decentralization and school-based management, where teacher performance is regularly measured and reported through student test scores, teacher evaluations, and other means, so that parents, teachers, and administrators can track school and student performance (Demas and Arcia 2015). Hence, convincing teachers to participate in CPD training in an environment where teacher accountability matters is bound to be relatively easy.

Financial incentives are the most common. They have an immediate effect on a teacher's household welfare and can be effective if designed properly. One common approach is pay-for-performance incentives, where student test scores are used as a metric to evaluate teacher performance (Glewwe, Ilias, and Kremer 2010; Muralidharan and Sundararaman 2011). Evidence from randomized controlled trials suggests that pay-for-performance incentives can help increase student test scores (Contreras and Rau 2012; Lavy 2002; Muralidharan 2012). Although financial incentives tied to teacher performance show some promise, administrators must be very careful to monitor teacher motivation and be vigilant against the potential negative effects of reduced collaboration among teachers. Leadership from the principal or head teacher, in these cases, is crucial for maintaining an appropriately professional climate in the school because teachers are very sensitive to imperfections in the school's approach to teacher accountability (Leithwood, Steinbach, and Jantzi 2002).

The share of motivation contributed by each factor and the ways in which a set of factors might work in concert to motivate teachers are not well understood. Implicit in the above list of incentives is the notion that each type of incentive will have a different impact at a different time, depending on how binding each factor might be at the time of its implementation. If teachers are paid relatively well, the impact of nonsalary factors could exceed that of salary factors. In other cases, higher pay for more specialized skills might make a difference. This is typically true for math teachers, where evidence shows that in countries where math knowledge among teachers is scarce, higher pay for math teachers may actually increase students' math skills (Carnoy et al. 2009).

2.2 Professional development programs and teacher motivation

Professional development programs give teachers an *a priori* opportunity to improve future performance. The literature on teacher motivation suggests that professional development programs should make certain that teacher motivation is addressed and monitored to ensure program success (Schieb and Karabenick 2011). An analysis of professional development programs in the United States indicated that teachers preferred professional training that improved their subject-matter knowledge, was fun to attend, and did not require too much time or effort (Karabenick and Conley 2011). These teachers' motivation to participate in professional training increased when their principals encouraged them to participate and if the training was in the form of a single workshop using a participatory method. Professional development programs that took time outside of the school year or that were lecture-based were not preferred by teachers.

Teachers' motivation to engage in CPD is important because empirical evidence suggests that CPD positively affects student learning. A meta-analysis of 16 studies on professional development reported that CPD training in math helped increase student test scores by an average of 0.21 standard deviations (SD), which is substantial (Blank and Alas 2009). Most of the studies in the meta-analysis concluded that the positive effect of CPD derived from teachers learning more about their specific subject content and improving their pedagogical skills, which helped them teach the content to students.

Most successful programs included follow-up reinforcement of learning, assistance with implementation, and additional support from more experienced teachers and mentors.

However, a key question was left unanswered: Were teachers motivated to implement what they learned in the program? It is possible that the discrepancy between the training content and the reality of the classroom makes it difficult for teachers to apply their newly acquired knowledge under the educational structure and managerial culture that they live in every day (Uysal 2012). Other research suggests that teachers' intrinsic motivation is the main factor influencing their decision to apply what they learned in a professional development program (Gorozidis and Papaioannou 2014).

Although little else is known about teachers' motivation to implement what they learned in a professional development program, recent evidence from Latin America strongly suggests that an education system's structure of incentives is the determining factor (Bruns and Luque 2014). An appropriate structure of incentives combines (1) a set of incentives that attracts intrinsically motivated people to the teaching profession, (2) school-based management that holds teachers accountable but directly recognizes teachers' effort and performance, and (3) a link between performance and financial incentives.

A key policy issue in education incentives is the need to link rewards to results. In this regard, *as long as teachers realize that their performance counts, then their willingness to participate in CPD programs and apply what they learned is bound to increase*. Several countries in Latin America have used a combination of incentives to improve teacher performance and student learning (Bruns and Luque, 2014; see also **Annex A**). The results are promising, but the policies require constant monitoring and refinement.

From the results obtained in Latin America, Bruns and Luque (2014) derived some guidelines for the design of incentives that could potentially be extended to Jordan:

- *Choose, and implement correctly, valid measures of teacher quality*. If criteria can be manipulated, they will be. Too much reliance on teachers' test scores may also reduce the number of teachers willing to be tested. Hence, the system must be calibrated over time.
- *Design tests carefully to ensure proper alignment with reward allocation*. The tests should cover teachers' content knowledge, pedagogical knowledge, and understanding of child development.
- Monetary incentives should be *temporary*, and the amounts should be reasonable *to increase motivation but avoid fraud*.
- *Teacher evaluators must be external to the system* to reduce social conflict or conflicts of interest.
- *Incentives must be based on political consensus* to ensure continuity.

In summary, a structure of teacher incentives must be in place for teachers to work in an environment that facilitates the implementation of innovations learned in professional development programs. This suggests that an incentive program, offering symbolic

professional recognition for classroom performance, may encourage teachers to implement practices learned in professional development.

2.3 Incentives for students

Generally, incentives designed to motivate parents and students tend to be mostly financial. For parents, there are two types of incentive structures: conditional cash transfers (CCTs) and food for education (FFE) programs. CCTs give money to poor households with school-age children contingent on their children's attending school on a regular basis. FFE programs can be of two types: those that serve breakfast and/or lunch to students every day, and those that provide take-home rations (THR), where students receive a certain amount of food staples, conditional on their maintaining a specified attendance rate. For students, monetary rewards are given when they reach a predetermined threshold of academic achievement.

CCTs and FFEs have proven successful in increasing student attendance, but in few studies have these programs had an impact on learning, underscoring the fact that improved learning likely depends on other factors besides attendance, such as teacher quality, the availability of good pedagogical materials, and other factors. Most CCT and FFE programs are implemented as components of poverty-reduction programs. As such, they tend to exclude the measuring and reporting of learning outcomes; poverty programs are generally implemented by agencies outside of the education sector. The evidence on learning outcomes tends to be *a posteriori,* generally showing mixed results.

2.3.1 Financial incentives

Financial incentives generally have a positive impact on student attendance and enrollment rates, as shown by their applications in Brazil (de Janvry et al., 2006), Cambodia (Barrera and Filmer 2013; Filmer and Schady 2011), Colombia (Attanasio, Fitzsimons et al. 2005; Barrera-Osorio et al. 2008), Costa Rica (Duryea and Morrison 2004), Ecuador (Schady and Araujo 2006), Honduras (Galiani and McEwan 2013), Jamaica (Levy and Ohls 2007), Malawi (Baird et al. 2011), Mexico (Attanasio, Meghir et al. 2005; Behrman et al. 2000; Schultz 2004), Nicaragua (Maluccio and Flores 2005), and Pakistan (Chaudhury and Parajuli 2010).

Although CCT and FFE programs could increase school enrollment and student attendance by as much as 26 percentage points (Filmer and Schady 2011; Meng and Ryan 2010), in most cases the impact of these incentives were more modest, increasing attendance rates by 3 to 10 percentage points. Paradoxically, in countries where the education sector has been successful in achieving universal coverage, the impact of CCT and FFE can be very small, as there are very few additional children to enroll.

2.3.2 Weak impacts on learning outcomes

Many CCT and FFE programs have no provisions for benchmarking learning outcomes, because the expectations seem to be that these types of incentives are aimed only at

increasing student enrollment and retention. Hence, most findings show increases in enrollment, attendance, and grade promotion. Levy and Ohls (2007), Duryea and Morrison (2004), Riccio et al. (2010), and Alderman et al. (2010) failed to find an impact of financial incentives on grade promotion in Jamaica, Costa Rica, New York, and Uganda, respectively. A few studies have found positive impacts on learning. In Nicaragua it was found that CCTs increased grade promotion by 7.3 percentage points (Maluccio and Flores 2005) and in Brazil by 6.2 percentage points (de Janvry et al. 2006). In Mexico, Behrman, Parker, and Todd (2009, 2011) found that the CCT program helped increase grade completion by half a grade for younger children, and close to one full grade for older children. In Bangladesh, the FFE program reduced dropout rates by 9 percentage points and students stayed in school up to 1.1 years longer (Meng and Ryan 2010).

CCT and FFE programs also had mixed impact on standardized test scores and course grades. Examining a CCT program for girls in Malawi, Baird et al. (2011) found weakly positive impact ranging from 0.12 to 0.17 SD on mathematics, reading comprehension, and cognitive ability tests. The Bangladeshi FFE program increased fifth-graders' achievement by 15.7 percent, but it also had a significant negative impact on the achievement scores of fourth-graders (Ahmed and del Ninno 2002). Evaluations of incentive programs in Jamaica (Levy and Ohls 2007), Ecuador (Ponce and Bedi 2010), Mexico (Behrman et al. 2000), and New York City (Riccio et al. 2010) also failed to find an impact on course grades or on standardized test scores in mathematics and language.

2.3.3 Pay for performance for students

The evidence from some experiments on student rewards suggests that well-targeted incentives that pay students for reading and for completing math assignments can have positive significant effects on achievement (Allan and Fryer 2011). The impact of financial incentives may be sustained, even after the removal of the incentives.

Under a randomized incentive program in India, treatment children were eligible to receive 100 rupees if they met their literacy goals. The impact was estimated to be 0.53 SD (Berry 2013). In Kenya, an incentive program provided scholarships to sixth-grade girls whose average test scores were within the top 15 percent of their school (Kremer et al. 2003). The scholarship program was estimated to have increased test scores across five core subjects by an average of 0.19 SD.

Angrist and Lavy (2009) examined the impact of a performance-based incentive program, in which low-achieving high-school students in Israel were eligible to receive cash payments for obtaining the high school matriculation certificate. These incentives increased certification rates by up to 8 percentage points in treatment schools. In Benin, Blimpo (2013) examined the impact of individual and group incentive structures for students in the 10th grade. Each of the incentive structures had a positive impact on achievement, with an effect size of 0.29 SD for the individual student incentives, 0.27 SD

for the team incentive, and 0.34 SD for the team incentive augmented with a tournament component.

In Cambodia, paid scholarships were offered to students on the basis of poverty or merit. Other than the criteria for selecting the recipients, both incentives were similar in terms of monetary value and in conditions for renewing the scholarship. The merit incentive structure showed an impact of 0.17 SD on the mathematics test and 0.15 SD on a digit span test, but the poverty incentive structure had no impact on either measure (Barrera-Osorio and Filmer 2013). The authors pointed out that there is an implicit trade-off in CCTs: Cash transfers to poor families induce attendance by poor students at schools that are not prepared to improve their learning outcomes. Using CCTs to target high-performing students may increase their learning, but given the clear link between poverty and low student performance, spending limited funds on targeting good students would reduce equity, since students from poor households likely would be left out.

Presumably, education is a desirable investment but, for children, this investment has a high discount rate—that is, any perceived benefits of education may be unknown to them or be too far into the future. Intuitively, educational incentives directed to children must have a simple internal logic: The value of the incentive has to be higher than the children's discount rate. Unfortunately, this discount rate is not well understood. Hence, implementing incentives to children becomes an empirical exercise where some incentives work and some do not.

2.3.4 Considerations for intrinsic motivation and sustainability

A concern of many policy makers is that paying students for academic performance could reduce their intrinsic motivation. However, multiple studies across contexts and age groups have failed to find a decrease in intrinsic motivation resulting from the financial incentives. Examining primary grade students in the United States, Bettinger (2012) failed to find any differences between the treatment and control students, in terms of either their own self-reports or teachers' ratings of students' intrinsic motivation. Similarly, neither Fryer (2011a) nor Fryer and Holden (2013) observed a decrease in intrinsic motivation over time, nor did they find any negative impact on a measure of self-reported "effort" index (e.g., complete homework, ask teachers for help). Studying sixth-grade girls in Kenya, Kremer et al. (2003) also found no decline in students' intrinsic motivation or attitudes toward school.

There is a notable dearth of studies that have examined whether the impacts from financial incentive programs are sustained after the removal of incentives. Three notable exceptions are Fryer and Holden (2013), Baird et al. (2011), and Kremer et al. (2003). Fryer and Holden (2013) found that two years after the termination of an incentive program that paid students for the number of mathematics objectives mastered, the treatment effect for high-achieving students remained, such that the impact on mathematics achievement was 0.271 SD. Baird et al. (2011) found that one term after a CCT program in Malawi had ended, the impact of the CCT program persisted, such that

the enrollment rate continued to exceed that of the control group. By contrast, the enrollment rates of students in an unconditional cash transfer program that was implemented and ended at the same time as the CCT program dropped to levels that were comparable to those of the control group. Similarly, Kremer et al. (2003) found that even one year after the incentives were removed, score gains remained relatively large, suggesting that the initial test score improvements reflected real learning

2.3.5 Larger incentives, larger impacts?

Although the topic is understudied within the literature, there is evidence of diminishing returns with respect to the size of the financial incentives, at least in international settings (see Allan and Fryer 2011 for a dissenting view using U.S. data). Filmer and Schady (2011) found that a US$45 scholarship in Cambodia had a very large impact on school attendance (i.e., 25 percentage points), but a US$60 scholarship did not significantly raise attendance above this level. Baird et al. (2011) varied the amounts of transfers to parents (between US$4 and US$10) or to girls (between US$1 and US$5), but found that increasing the transfer amounts had no effect on any of the outcomes for the CCT treatment. In addition, Levitt et al. (2013) found that younger children (i.e., second and third graders) in the U.S. did not respond differentially to low- and high-financial rewards (US$10 and US$20, respectively).

2.3.6 Unintended consequences

Despite the generally positive impacts of financial incentives on school enrollment and attendance, Barrera-Osorio et al. (2008) found evidence of a reallocation effect, such that siblings of students who were eligible for cash awards were less likely to attend school and more likely to go to work. Namely, when they compared households that registered two children, they found evidence of lower school attendance and more labor market work for an untreated child with a treated sibling compared to an untreated child with a similarly untreated sibling. In this case, the interplay between education and poverty shows that the income needs of a family may themselves be an incentive, pushing for short-term income over the long-term financial benefits of education.

Similarly, Fryer and Holden (2013) described a substitution effect such that subject areas that were not incentivized showed decreases in performance, relative to the control group. In Houston, where the number of mathematics objectives mastered was incentivized, Fryer and Holden (2013) observed an increase in mathematics achievement on the state achievement test. However, they also observed a decrease of 0.078 SD on the state reading achievement test. The negative impact on reading achievement persisted for lower-achieving students, even two years after the removal of the incentive program.

3. Previous work on incentives in Jordan

A teacher motivation study, also conducted under the Education Data for Decision Making (EdData II) project,[1] attempted to investigate the issue of teacher motivation in general, to ascertain a way to increase teachers' willingness to implement effective teaching practice. The current study is motivated by the findings of the *Intervention Impact Analysis Report* (Brombacher et al. 2014) under the same task order. Specifically, with regard to teacher motivation, it was found in the 2014 intervention impact study data that, although teachers recognized and affirmed the intervention's positive impact on student performance and their teaching practice, they were reluctant to express support of the continuation of the program. Indeed, more than three out of four teachers indicated that they would not want to continue the program if given the choice. The intervention impact report could only speculate at the underlying reasons for this particular finding, and the teacher motivation study presented an opportunity to be able to understand more fully what motivated and demotivated teachers about their profession, as well as how a continuing professional development program might be designed that would encourage teachers to participate in professional development and implement in their classrooms what they learned.

Data were collected in two collection phases, through two instruments (one per phase): (1) a focus group discussion conducted with 41 teachers and (2) a 29-item questionnaire administered to 37 teachers.

During the focus group discussions, most male and female teachers asserted that they became teachers because they thought that they would like the job, or because they had always wanted to be a teacher. For what motivates them about their profession, most teachers expressed that appreciation (from colleagues or supervisors) or support from head teachers was motivating, more so than pleasant classrooms or extra resources. Rather than perceiving the physical school and classroom environment as being a motivating factor, however, teachers saw it as a demotivating factor. Teachers mentioned the number of students and the condition of schools and classrooms in this way.

For professional development, teachers expressed appreciation for programs that responded directly to their needs as professionals and for those programs that were conveniently timed. Most teachers suggested that they would be more motivated by professional development were it more directly tied to supervisory appraisal and rankings among colleagues. Recognition and compensation were noted as factors that might encourage further participation in professional development.

When asked about what motivated them as professionals, teachers mentioned student performance and support and encouragement from their head teachers. Aspects of

[1] EGRA/EGMA/SSME National Survey in Jordan, Task Order No. AID-278-BC-11-00019 (RTI Task 16), December 2011–November 2015.

teaching that demotivated them, however, were lack of parental support, lack of teaching resources, and lack of support from supervisory staff.

When asked to rank two sets of five motivating factors, teachers highly ranked an attractive classroom and school environment, as well as student performance (as opposed to other factors such as prestige, teaching resources, support from supervisors, or parental appreciation). When asked to rank two sets of five potentially demotivating factors about their profession, teachers most often mentioned student performance (presumably poor performance) and crowded classrooms.

When asked to rank two sets of five reasons for attending in-service training, teachers highly ranked improving student performance, increasing teacher subject knowledge, and the relevance of the course content to their work. When ranking features of in-service training that were important to them, teachers mentioned the quality of the course content. When ranking factors that motivated them to implement professional development learning, teachers highly ranked wanting to improve student performance.

For benefits (i.e., incentives) that a continuing professional development program could offer participating teachers, respondents expressed interest in promotion and in salaries tied to program participation.

The analysis of these data suggests that while many factors contribute to teachers' motivation in Jordan and, in particular, their motivation to participate in professional development activities, reward (particularly financial) is key. A continuing professional development program that linked the participation in and implementation of in-service teacher training activities to salary, promotion, and appointment opportunities would be worthy of exploration.

4. Pilot program description and implementation

This pilot incentive program was designed and implemented in parallel with the Reading and Math Program (RAMP), a 5-year program in Jordan for the early grades, sponsored by the United States Agency for International Development (USAID). The incentives were designed to align with the first type of incentives discussed in the literature review above (i.e., professional, or symbolic, rewards). Rather than providing cash or token rewards to students or teachers, the program offered symbolic incentives in the form of a certificate and public recognition at a local awards ceremony for successful participation in the program. The pilot extended incentive programs to students as well as teachers, in hopes of encouraging the students to read more often at home—an objective of RAMP—and incentivizing the teachers to put into practice what they had been trained to do as part of RAMP. The incentive programs for students and teachers are described in further detail below.

5. Incentive programs

5.1 Grade 2 students

The student incentive pilot (*n* = 47 schools) tracked student and household engagement in reading activities outside of school. Grade 2 students were told that if they read at home on a regular basis (at least three days per week), completed a provided reading log with their parents/guardians, and returned the log to school, they would receive recognition at an award ceremony at the end of term and certificate. Students and parents were informed of the program by principals, who were in turn informed by RAMP coaches. Students were given a reading log. Students and their families received specific instructions on how to fill out the reading logs.

The goal of this program was to determine whether incentives in the form of formal acknowledgment of student reading practice would increase the frequency of independent reading at home. RAMP coaches collected the information during coaching visits throughout the 8-week implementation period, and gave it to the RAMP data collection team at the end of the term.

5.2 Teachers

The teacher incentive program (in the same 47 treatment schools as the student program)[2] tracked teachers' implementation of RAMP. Teachers were informed that if they exhibited effective implementation of the classroom teaching practices they were trained to use, they would receive special recognition in a ceremony at the end of term. This recognition would be in the form of a certificate if they received a teacher implementation score of 4 or higher (on a 5-point scale) for the majority of their observed lessons. Classroom observation data, which were already being collected by coaches on a near twice monthly basis as part of RAMP, were used to evaluate the teacher implementation in the classroom. In the observational protocol, data were collected on teachers' use of teaching materials, their use of diagnostic tools (and assessments), and effective teaching practices. These observations were then aggregated to determine a composite "implementation score" for each teacher on each visit. The purpose of this pilot was to determine whether effective implementation of RAMP could be encouraged by the formal recognition of high-quality work undertaken by teachers.

Teachers were observed between two and six times during the semester; not all teachers received the same number of coaching visits (and, thus, observations). The number of visits was not under the control of the pilot program; rather, it was determined by the RAMP coaches. During the observations, coaches collected data on teaching practices described above. These data were then transmitted to the RAMP coaching database.

[2] The two treatments were conducted at the same schools as they were determined to be independent from one another and therefore not likely to influence each other's outcome.

5.3 Control schools

Control schools (*n* = 42 schools) did not participate in the student or teacher incentive programs, in that students and teachers in these schools were not told that they would be eligible for awards if they met the program criteria. Indeed, they were not made aware of the incentive program at all, but were given information that was otherwise identical to that of their counterparts in treatment schools. Students in control schools were also encouraged to read at home, fill out the reading logs, and return them to school. These data were collected by RAMP coaches during classroom visits. Classroom implementation scores were also calculated for teachers in control schools during coaching observations.

6. Communication plan

The treatment and control portions of the program were communicated by different printed fliers delivered to schools by RAMP coaches. In addition, RAMP coaches disseminated information about the program to principals, grade 2 teachers, students, and parents in a cascade model (that is, RAMP staff first informed coaches, who then informed others).

The materials that the RAMP coaches provided to pilot program schools are as follows:

- **Fliers** (advertising materials) that stated information about the different programs (student and teacher incentives). Fliers generally were directed to one actor (students, teachers, or parents). As noted, the fliers also differed for treatment and control schools.
- **Reading logs** for the students to fill out as they read at home.
- **Teacher recording logs** for teachers to keep track of student reading on a weekly basis. Teachers recorded the number of books students read for the week and showed these logs to coaches during their classroom visits.

7. Sampling methodology and school selection

Participating schools for this pilot study were selected through a purposeful, iterative process. Attempts were made to randomize certain aspects of the selection process; complete randomization, however, was not feasible for this study.

The Jerash governorate was selected as the geographic location for this study as it is relatively compact and the number of schools (approximately 100) was manageable for a pilot with several treatment conditions. Of 110 schools in Jerash, 21 schools were served only by Ministry of Education coaches (as opposed to RAMP coaches) and were therefore not selected to participate in the treatment. Randomization would have been ideal for sorting the remaining 89 schools into treatment and control groups, but there was concern that schools positioned in close proximity to one another might induce program (or at least information) leakage from treatment to control schools. In response

to this, schools were grouped into geographic clusters; these clusters were then randomly selected into the treatment (47 schools) and control (42 schools) groups.

Within these schools, all grade 2 teachers and students participated in the incentive program. Students, however, were not forced to participate in the program; during program communication, it was stressed that participation was voluntary and would not count in one way or another toward their school grades.

Similar to the Brombacher et al. (2014) study, the incentive program used focus groups to collect information directly from participants—in this case, both teachers and students. A small number of grade 2 students (n = 11, 6 females and 5 males) and teachers (n = 13, all female) from three treatment schools were selected to participate. The selection of participants for focus groups was not randomized. Three focus groups were held for teachers and three were held for students.

8. Data collection

For the student incentives, data on student reading were first recorded by students in their reading logs, then tabulated by teachers at the classroom level. Each week, students would return their reading logs to school and teachers would record how many books the participating children read each week for the 8 weeks of implementation. Baseline data on student reading were not collected prior to the intervention.

For the teacher incentives, coaches observed teachers' classroom instructional practices, specifically to determine the extent to which teachers were implementing RAMP. Coaches collected data on the following teacher practices: teachers' use of teaching materials, their use of diagnostic tools (and assessments), and effective teaching practices. These data were entered electronically into observation software on tablets and uploaded into the RAMP coaching database. Baseline data were collected on classroom teaching before the intervention.

The focus groups were facilitated in Arabic by RAMP staff with the participating teachers and students from the three selected schools. Discussions were recorded during the focus groups, and afterward the audio was transcribed in Arabic and translated into English.

9. Results and discussion

9.1 Student incentive pilot

The student pilot was conducted with 1,244 male students (751 in control schools and 493 in treatment schools) and 1,444 female students (628 in control schools and 816 in treatment schools). Female students constituted 46% of the sample in control schools and 62% of the sample in treatment schools. **Table 1** tabulates additional descriptive statistics of control and treatment schools: total number of books read per student over

the 8-week pilot, the average number of books read per week, the proportion of students who met the incentive criteria (i.e., who read at least three books per week, or 24 books over 8 weeks), as well as those who managed to read four, five, and six books, on average, per week, and a composite school quality variable ("teacher quality").

Table 1. Student incentive pilot descriptive statistics

	Control		Treatment	
	Average	SE	Average	SE
Male (*n* = 1,244)	54.5% (*n* = 751)	—	37.7% (*n* = 493)	—
Female (*n* = 1,444)	45.5% (*n* = 628)	—	62.3% (*n* = 816)	—
Total books read	14.0	0.36	19.2	0.42
Average books/week	1.8	0.04	2.4	0.05
Students met criteria (3 books per week)	29.0%	1.2%	32.7%	1.5%
4 books per week	17.7%	0.7%	17.5%	1.1%
5 books per week	8.4%	0.7%	10.8%	1.1%
6 books per week	3.5%	0.6%	5.4%	0.8%
Teacher quality	23.6%	5.8%	33.9%	6.4%

SE = standard error.

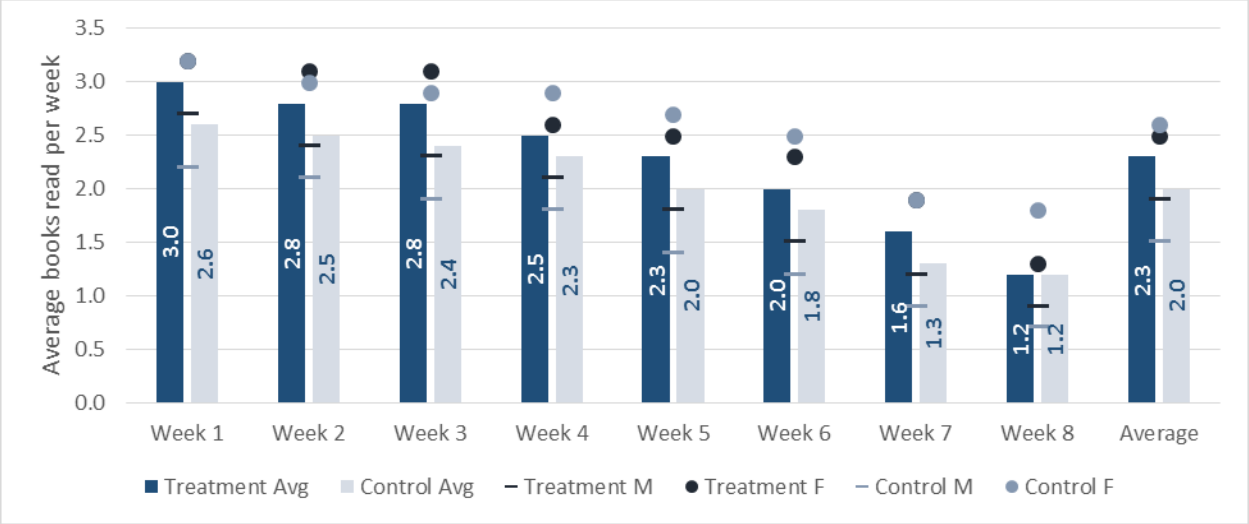
As seen in the table, more students met the criteria in treatment schools (33%) than in control schools (29%). Similar results were seen in most other benchmarks (with the exception of the proportion of students who averaged four books per week, which was virtually identical between treatment and control schools): slightly more students in treatment schools read five books (11% as compared to 8% in control schools) or six books (5% as compared to 4% in control schools) than peers who were not offered the incentive. We explore below which of these differences were statistically significant.

The teacher quality variable (minimum of 0.0, maximum of 1.0) represents the proportion of teachers in the school who met the incentive criteria for teachers at baseline (i.e., these teachers received a score of 4 or higher on the majority of their classroom observations before the introduction of the teacher and student incentives). This serves as a rough proxy for the quality of classroom instruction in subsequent analyses. More teachers in treatment schools (34%) met the incentive criteria at baseline as compared to control schools (24%). To the extent that this variable captures information about the quality of teaching practices in schools, it may signal that instructional quality was already slightly higher in treatment schools than in control schools. This in turn brings into question whether schools were indeed randomly selected into treatment and control

groups during the allocation process; if not, this may have biased the results. Nonrandom allocation to groups in a study such as this one may be related to the amount of resources a school has at its disposal (e.g., more books for reading), and the level of support teachers give students as they learn to read: better teachers may self-select into higher-functioning, well-resourced schools. As such, the following analyses should, to the extent possible, control for this potential for bias.

Figure 1 displays the mean number of books read in treatment and control schools for each of the 8 weeks of the pilot, as well as the 8-week average. These statistics are also disaggregated by sex. Most weeks show a differential between treatment and control schools vis-à-vis the average number of books read per week: More books were read in treatment schools during all weeks save the eighth. The figure also makes clear that, while girls read more books per week than boys on average, the treatment–control difference was due mainly to boys' participation. There was not a distinct pattern in terms of girls' reading with regard to treatment and control schools, but boys in treatment schools tended to read more than boys in control schools throughout the 8-week pilot. In other words, the treatment may have had an impact on the frequency with which grade 2 boys read at home.

Figure 1. Average books read per week in treatment and control schools, by sex



From the figure, it is also evident that all children read less frequently from week to week: Both boys and girls showed marked declines in reading frequency from Week 1 to Week 8 of the pilot. Indeed, the average number of books read at the end of the pilot in both treatment and control schools was less than one-half of what it was at the beginning. As this phenomenon occurred in both treatment and control schools, it is unlikely to have been due solely to a decline in effectiveness of the treatment (i.e., in this case, students might have lost interest over time in the symbolic incentive on offer). Rather, the focus group data suggested a number of factors unrelated to the pilot: the onset of Ramadan during the end of the pilot, a focus on end-of-term exams, and

teachers' feelings of being overburdened or disinterested in the pilot. In focus groups, some teachers indicated that the pilot program was not supported by all parents. In other words, they as teachers had to follow up with some of their students ("weak students" in particular) in order for students to return the reading log.

These findings do suggest that, while improvements were seen between treatment and control schools in all but the last week of the pilot, a symbolic incentive to read with greater frequency may not be powerful enough to overcome systemic influences on students' reading habits, such as sociocultural factors, school calendars, and parent or teacher disinterest.

In addition, several teachers from focus groups mentioned that resources necessary for the successful implementation of the pilot were lost over the course of 8 weeks. In at least some cases, stories that were lent to students to take home and read with parents were not returned to school, and therefore students didn't have anything to read in the last weeks of the pilot. Some students also lost their reading logs, or stopped returning them to school. While issues such as missing books and reading logs are not insur-mountable, this pilot did not include continuous support and monitoring of resources.

Table 2 tests whether we can be confident that the patterns described above were not products of chance. Two-sample *t*-tests were used to compare differences between treatment and control schools in terms of the proportion of students who met the incentive criteria (three books read per week), as well as four, five, and six books per week. Asterisks indicate statistically significant differences between treatment and control schools in the subpopulation of interest.

Table 2. Proportion of students meeting incentive criteria

		3 books (criteria)		4 books		5 books		6 books	
Group	Sex	*n*	%	*n*	%	*n*	%	*n*	%
Control	Male	125	16.6	65	8.7	38	5.3	13	1.7
	Female	278	44.3	182	29.0	67	13.1	36	5.7
	Average	405	29.0	247	17.7	105	8.4	49	3.5
Treatment	Male	123	25.0***	70	14.2**	41	8.8**	22	4.5**
	Female	305	37.4**	159	19.5***	89	11.9	48	5.9
	Average	430	32.7*	230	17.5	131	10.8*	71	5.4**
Total	Male	248	19.9	135	10.9	79	6.7	35	2.8
	Female	583	40.4	341	23.6	156	12.4	84	5.8
	Total	835	30.8	477	17.6	236	9.6	120	4.4

*$p < 0.05$; **$p > 0.01$; ***$p > 0.001$.

Tests confirmed statistically significant differences between the proportion of all students who met the incentive criteria (33% in treatment schools versus 29% in control), who read five books (11% in treatment schools versus 8% in control schools), and who read six books per week (5% in treatment schools versus 4% in control schools). When we disaggregated by sex, more males in treatment schools met the incentive criteria and higher benchmarks than in control schools. These differences were statistically significant.

Statistically significant differences were also found between female students in treatment and control schools as well, although in the opposite direction. A larger proportion of female students in control schools met the incentive criteria (although the incentive was not offered at control schools) and most of the higher benchmarks. Only the proportions of female students meeting the incentive criteria and reading four books per week were found to be significantly different, however. These results suggest that the offer of a symbolic incentive may have encouraged male students to read with increased frequency, while the incentive may not have had the same effect on female students.

Importantly, these results were replicated when we used weekly reading data (rather than 8-week averages): Even as the frequency of reading decreased, the association between a symbolic incentive and out-of-school reading remained significant overall, and for male students. Focus groups questions did not probe this potential difference between boy and girl students, and teachers did not suggest that boys were more motivated by the incentive than girls; rather, interviewed teachers generalized that all students were interested in the program.

We also tested to determine whether these findings were a product of the relative averages for male and female students (i.e., because females tended to read more than males in all schools). When we controlled for overall reading level—limiting the comparison to those students who averaged less than three books per week—our findings were reversed: Among infrequent readers (i.e., those students who averaged less than three books per week during the pilot), girls in treatment schools tended to read more than girls in control schools. While this difference was significant ($p < .001$), it was not substantial (girls in treatment schools read 0.2 books per week more than girls in treatment schools). A statistically significant difference was not detected among boys who infrequently read. As such, it does not appear that gains in reading frequency seen in treatment schools (relative to control schools) were a product of lower base levels of reading in certain populations.

To determine whether the differences in average books read per week were robust to several background variables, we created a simple linear regression model that controlled for student sex, the class the student was in, and teaching quality at baseline (derived from the teacher observation baseline). This simple model allowed us to see if the association between treatment and more frequent reading at home would hold when we controlled for student sex, interclass differences, and the quality of classroom

instruction. Four models were fitted to the data: an overall model, one for male students, one for female students, and one for infrequent readers. Results are shown in **Table 3.**

Table 3. Relationship between symbolic incentives and average books read during the pilot

	Overall		Males		Females		Low readers	
	Coef	SE	Coef	SE	Coef	SE	Coef	SE
Treatment	0.69***	0.07	0.68***	0.09	0.44***	0.10	0.27***	0.04
Teacher quality	-0.30**	0.09	-0.37**	0.12	-0.64***	0.13	0.10*	0.05
Class	-0.08*	0.03	-0.07	0.043	-0.04	0.05	0.00	0.02
Female	0.00	0.00	—	—	—	—	0.00	0.00
Constant	1.95***	0.07	1.50***	0.09	2.58***	0.11	1.04	0.04
R^2	0.05		0.06		0.04		0.04	

*p < 0.05; **p > 0.01; ***p > 0.001.
SE = standard error.

The table shows the treatment as being consistently associated with an increased frequency of reading at home across the four models, even when we controlled for sex, class, and instructional quality. The treatment was, on average, associated with an increase in 0.7 books read per week. The association between treatment and increased reading was similar in the male model (0.7 more books per week), but weaker in the female model (0.4) and the model for low readers (0.3 more books per week). The class variable was significant in the overall model, but that significance disappeared in the male, female, and low reading models.

The teacher quality variable also was significantly associated with reading, but the directionality (i.e., negative) was different from what we expected. Lower proportions of teachers meeting the incentive criteria at baseline were associated with increased frequency of reading outside of school, particularly for female students (0.6 books per week). This negative association between reading and teacher quality negates the potential bias discussed above with regard to this variable. More teachers in treatment schools (33%) met the incentive criteria at baseline than in control schools (25%); thus, we were concerned that higher quality instruction might be positively related to student reading and therefore would bias any treatment effects seen. This model suggests, however, that teacher instructional quality (as defined by classroom observation scores) was not positively related to an increased frequency in student reading outside of school.

This may have been the case for a number of reasons, but perhaps the foremost is that the teacher quality variable measured in-class behaviors of teachers and not the out-of-school behavior of students. Teachers with higher observational scores may have been providing more time for their students to read during the school day and may therefore

have been less likely to encourage them to read at home. It is also the case that teachers received higher observational scores for fidelity in implementing RAMP; perhaps lower quality teachers were more likely to implement the program with fidelity than were higher quality teachers. Such teachers may have received higher observational scores, but may not have encouraged their students to read outside of the classroom. These rationales are speculative; suffice it to say here that teaching quality as measured by observation-based scores does not positively predict the frequency of out-of-school reading.

Focus group data also revealed a selection of students' and teachers' perspectives on the student incentive program. While these are not representative of all students and teachers, they are illustrative of their respective experiences with the program. For their part, students interviewed for this report indicated that they, for the most part, enjoyed reading and that the pilot program was therefore interesting to them. Focus group participants suggested that they enjoyed what they saw as the competitive nature of the program: Because all students could see how many books their classmates read on a weekly basis through the teacher's log, they attempted to read more books than their friends. Although this was not the intention of the pilot program, the sense of competition appeared to motivate grade 2 students.

Teachers who participated in focus groups also claimed that the pilot incentive program was motivating to their students and that the students enjoyed competing with each other. Interestingly, students' interest in the program did not appear (according to interviewed teachers) to be limited to strong readers; rather, most students were motivated to read stories, record them in their reading logs, and report back to the teacher or their fellow students. Numerous teachers suggested that their students, even students who were not "frequent readers," developed a habit of reading.

> *"I would apply the initiative again because I have some very weak children who improved significantly. They even compete with one another now and rehearse at home prior to reading in class. This supported their vocabulary and made them better [readers]." (T2FG3)*[3]

> *"When I read them a story before, there was no particular reaction. I asked them questions like: 'Did you read the story?' They would reply, 'No, we didn't.' Later I explained the initiative to them and Dr. Sharifa [the principal] gave them these reading logs. There was competition among them for reading stories…I gave them stories to read and report to me in the next day, and they would read the entire story and tell the rest of [sic] students about it. All students want stories now. Every student now buys stories, reads them, and tells the rest of the class about them." (T4FG3)*

> *"One of my students was very weak, but reading stories helped her improve significantly: her vocabulary expanded and her thoughts*

[3] Teachers' quotes are labeled by number and focus group. Thus, T1FG1 refers to teacher number 1 in the focus group at school 1.

improved. She was a bit frustrated but when I encouraged her to read stories, she loved to study and her achievement improved." (T2FG2)

"Yes, they read more. A student of mine takes more and more stories each week. Everything about him changed: his information, his thoughts, and his values. They all changed. It was really lovely. A quantum leap." (T3FG1)

Further, interest in the program did not seem to be limited only to elements of the program (i.e., reading stories and filling out the reading log). Several teachers who participated in focus groups told of students who routinely reported what they learned from the books:

"My students would tell me about any new information they acquire from reading. A student wouldn't [sic] imagine that whales sleep. One day a student came to me and said, 'Look! Whales fall asleep!' She thought only human beings sleep." T5FG1

Others noted that several of their students had begun creating and writing stories of their own to read and share.

While students and teachers, on the whole, perceived that the pilot program had had a positive impact on the frequency of student reading at home, teachers did mention several negative aspects of the program. First, some teachers noted that not all parents supported the program. They suggested that not all parents helped their children fill out the reading log, or did not return the log or the reading stories to school. Other teachers interviewed for this study cited their other burdens (e.g., trainings associated with RAMP, leading extracurricular activities) which rendered devoting attention to this pilot difficult. Teachers perceived the incentive to be inequitable with regard to weak readers: Some focus group participants thought the incentive criteria to be unattainable for the weakest readers, who could read, in their estimation, "maybe two to three pages" per night.

In summary, data collected for the student pilot showed that the offer of a symbolic incentive may have encouraged grade 2 boys to read more frequently at home, but the same may not have been true of grade 2 girls. These findings held true even when aggregate reading frequency was low (i.e., at the end of the term) and when we controlled for possible class effects and teaching quality. While we were not able to collect baseline data against which to compare our findings, the quantitative analysis suggests that symbolic incentives may be a means to encourage boys to read more frequently outside of school, a practice that is associated with better overall reading performance and comprehension. Focus group data from students and teachers revealed that students were motivated by the incentive pilot program and that it may have cultivated new habits of reading in some students. Teachers interviewed for this study offered descriptions of numerous students who improved or changed their habits as a result of the pilot (as perceived by the teacher). There was no indication in focus group data that boys were more likely than girls to be motivated by the offer of a symbolic incentive.

9.2 Teacher incentive pilot

The symbolic incentives for teachers rewarded successful implementation of the RAMP program, as determined by scores from observed classroom lessons. As mentioned previously, baseline data were collected on teaching practices prior to the incentive pilot; however, basic demographic data were not available. **Table 4** presents what descriptive statistics were available for the teacher pilot: the number of coaching visits, teachers' average score, and whether they met the incentive criteria (i.e., whether they received a score of four or higher in the majority of their coaching visits). These data are displayed for control and treatment schools both before and after treatment. Statistically significant differences between control and treatment schools are denoted by asterisks.

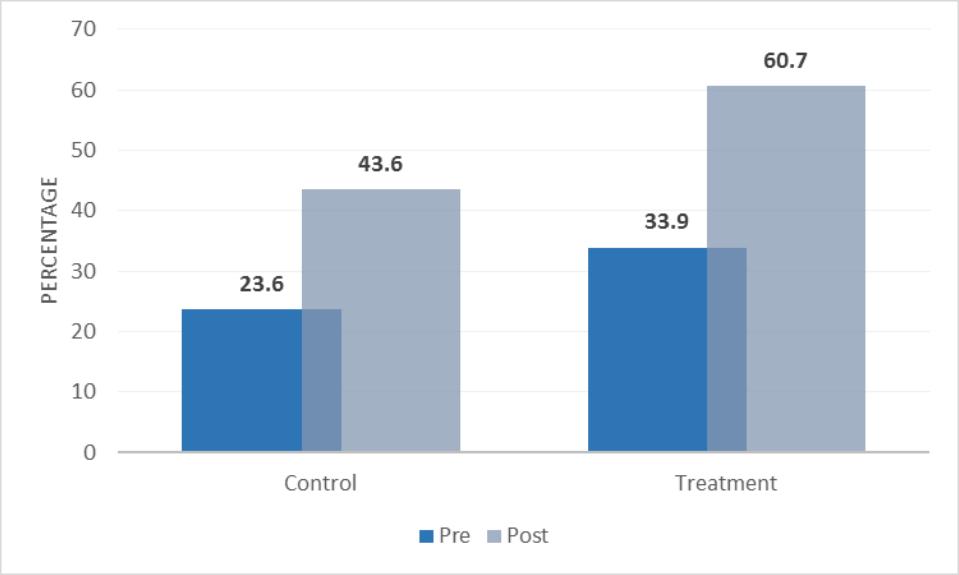
Table 4. Descriptive teacher statistics

		Control		Treatment	
Timing	Statistic	Average	SE	Average	SE
Pre-treatment	Number of visits	2.0	0.1	2.2	0.1
	Average score	2.4	0.1	2.7*	0.1
	Met criteria	23.6%	5.8%	33.9%	6.4%
Post-treatment	Number of visits	3.3	0.1	3.5	0.1
	Avg score	3.0	0.1	3.2*	0.1
	Met criteria	43.6%	6.7%	60.7%*	6.6%

*$p < 0.05$.
SE = standard error.

Teachers in treatment and control schools received virtually the same number of coaching visits before and after treatment. The average observation score differed significantly both before and after the onset of the incentive activity, with teachers in treatment schools receiving higher scores, on average. This reflects the teaching quality variable discussed in the previous section: More teachers met the teacher incentive criteria in treatment schools at baseline. The proportion of teachers meeting the incentive criteria differed significantly after treatment: 61% of teachers in treatment schools met the criteria as compared to 44% of teachers in control schools. **Figure 2** graphically displays the proportion of teachers who met the incentive criteria in control and treatment schools before and after treatment.

Figure 2. Percentage of teachers meeting incentive criteria, pre- and post-treatment



From Figure 2 and Table 4, it is clear that while treatment schools at baseline had a larger proportion of teachers who met the incentive criteria (although the difference was not statistically significant, due to high standard errors), the difference between treatment and control schools after treatment was larger (61% of teachers met the criteria in treatment schools as compared to 44% in control schools—a difference of 17 percentage points).

Since increases were seen in both control and treatment schools, we created two regression models to determine whether teacher performance (as measured by classroom observation scores) was associated with the offer of a symbolic incentive when we controlled for other variables. We created a linear regression of teachers' average post-treatment observation scores and a logistic regression of whether teachers met the incentive criteria. Both models attempted to test whether the association between the treatment and teachers' observation scores was robust to controlling for the number of visits received by teachers and teachers' baseline performance. We controlled for these factors because the number of visits from coaches may have influenced a teachers' ability to improve and therefore implement the program with fidelity and because teachers' classroom teaching during the pilot is likely to have been influenced by prior teaching performance. Results are shown in **Table 5.**

Table 5. Association between treatment and classroom observation scores

Linear regression – Avg observation score			Logistic regression – Teacher met criteria		
Variables	**Coef**	**SE**	**Variables**	**Odds ratio**	**SE**
Treatment	0.05	0.14	Treatment	1.79	0.75
Number of visits	0.02	0.09	Number of visits	1.13	0.31
Pre avg score	0.66***	0.10	Pre met criteria	6.18***	3.17
Constant	1.30**	0.40	Constant	0.34	0.33
$R^2 = 0.312$			**Pseudo $R^2 = 0.120$**		

*$p < 0.05$; **$p > 0.01$; ***$p > 0.001$.
SE = standard error.

Results from both models indicate that the association between a symbolic incentive and teachers' classroom observation scores was not significant when we controlled for teachers' performance prior to treatment. In the models, neither the treatment nor the number of coaching visits received by the teacher was associated with average observation scores or the likelihood of meeting the incentive criteria. Only the prior performance variable had a statistically significant coefficient.

In the linear regression, a one-point increase in teachers' prior average observation score was associated with a 0.7-point increase in their post-treatment average score. The logistic regression shows that teachers who met the incentive criteria prior to treatment were more than six times more likely to meet the incentive criteria after treatment. Interestingly, this finding was replicated when we reran the model without the prior performance variables: Neither the treatment nor the number of visits was significantly predictive of teachers' observational scores. In other words, the offer of a symbolic incentive does not appear to have been associated with enhanced teaching performance (as observed by coaches).

While this seems counterintuitive given the larger nominal increase in the proportion of teachers meeting the incentive criteria in treatment schools (27 percentage point increase) as compared to treatment schools (20 percentage point increase), these nominal figures represent an increase of 85% in control schools and 79% in treatment schools. The regression models, as simple as they are, indicate that we cannot be confident that the treatment (a symbolic incentive) was associated with enhanced teaching practices. Why might this not be so? It is likely because the incentive criteria were inherently tied to RAMP: Teachers would receive higher observational scores (and therefore become eligible for symbolic incentives) if they implemented the program well. Implementation with fidelity undoubtedly was related to exposure to the program—the amount of time teachers had to understand the program and what was expected of them—and both treatment and control schools were exposed to RAMP for the same amount of time. As such, it is likely that the growth in the proportion of teachers who met

the incentive criteria in both treatment and control schools was a product of prior teaching performance and program exposure, as well as other teacher and school-level variables (e.g., training, experience).

To conduct a more robust assessment of the treatment on teacher observation scores, we also created and tested a difference-in-difference model. This approach attempted to determine whether the difference in the growth in teacher observation scores seen in treatment schools was significantly different from that observed in control schools, while controlling for differences in baseline teacher observation scores.

When we used teachers' post-treatment average scores as the outcome variable in a linear regression model, the difference-in-difference effect was marginally non-significant ($p = 0.07$). A logistic regression model (with the outcome variable as whether or not teachers met the incentive criteria) also showed similar results: The difference-in-difference effect was not statistically significant ($p = 0.81$). In sum, while there was indeed a nominal difference between treatment and control schools in terms of the proportion of teachers who met the incentive criteria, this difference disappeared when we controlled for teaching performance at baseline. What growth we saw in treatment schools in teacher observation scores, therefore, likely was similar to what we would expect to see even if the incentive were not offered.

Although the quantitative analysis presented above did not find that teachers in treatment schools were more likely to experience gains in observation scores from baseline to the end of the pilot, the focus group discussions suggested that teachers interviewed for this study were indeed motivated by a symbolic incentive. Overall, teachers expressed support for a formal certificate based on performance, particularly if it was presented officially and publicly (i.e., in front of colleagues, the principal, and the community). They noted that such recognition tended to increase their motivation and gave them the impression that "someone cares" about the effort they put forth. One teacher suggested that symbolic incentives like those on offer could heighten perceptions of teaching as a profession in Jordan. It was clear in focus groups that the teachers desired recognition for the work they conducted with students and for implementing education initiatives (new programs, curricula, etc.). Aside from symbolic incentives, teachers interviewed for this study noted that the following would motivate them in their profession: financial rewards, celebrations, community recognition, and a reduction in other professional "burdens."

In summary, the offer of a symbolic incentive was not found to be associated with enhanced teaching practices, as measured by classroom observation scores. While growth in the proportion of teachers who met the incentive criteria was observed in treatment schools (34% to 61% of teachers), growth was also seen in control schools (24% to 44% of teachers). Our models indicated that this growth could not be confidently associated with the treatment. This is likely because performance against the incentive criteria may have been a product of prior teaching ability and program exposure, neither of which can be influenced by a performance incentive. However, this does not imply

that symbolic incentives do not motivate teachers in Jordan. Indeed, the focus group data suggest that teachers interviewed for this study did find formal certificates and public acknowledgement to be deeply motivating, as they reflect formal recognition of effort and performance. Instead, this pilot may have been too short-lived to have provided evidence of more advanced classroom instruction in treatment schools; a longer implementation period may have been needed.

10. Limitations

A number of limitations decreased the confidence we have in these findings. These are discussed below.

10.1 Lack of randomization

There is evidence that the allocation of schools into treatment and control groups was not random: There were significantly more teachers in treatment schools who met the incentive criteria at baseline than in control schools. Recall that schools were first placed into geographical clusters to avoid information leakage from treatment to control schools. These clusters were then randomly allocated to treatment and control groups. Evidence of systematic differences between these groups at baseline, however, suggests that potentially influential characteristics (such as student composition, teacher quality) were not randomly distributed between treatment and control schools. This is not necessarily because the allocation was conducted improperly, but rather because schools that were clustered together were more likely to be similar (and to serve similar demographics) than other schools.

10.2 Lack of background variables (including a baseline for student reading)

This pilot lacked background variables for students, teachers, and schools that would normally be used to control for exogenous factors that could be related to the frequency of student reading (e.g., SES) and teacher observation ratings (e.g. training, experience). As such, the above analyses did not ascertain whether the significant findings were influenced by related, but unmeasured, variables. Subsequent iterations of a symbolic incentive program (and its evaluation) should make sure to collect basic background variables for students (e.g., sex, age), teachers (e.g., sex, experience, training), and schools (e.g., urbanicity). Ideally, background information should also include baseline data for students and teachers. While this pilot study included baseline information on teacher observation scores, it did not include baseline reading frequency for students, which should be made a priority in future studies.

10.3 Short implementation (made shorter by some schools stopping, competing programs, resources not brought back to school)

The implementation period for this pilot was only a short 8 weeks due to longer-than-expected start-up times. As such, a number of compromises were made in the design and implementation process. For example, we considered having two assemblies for

students and teachers who met the incentive criteria, since a school term is a relatively long time to wait with no performance feedback (vis-à-vis the incentive). A midterm assembly might also have increased interest in the incentive from other students and teachers. However, time limitations rendered this an impossibility. The compressed implementation time also made the treatment effect more susceptible to short-term shocks, such as book shortages, absenteeism, and competing school programs. Short implementation time also may have led some teachers and parents to view this pilot as unimportant.

10.4 Communication plan may lead to variation in how the program was presented, communicated, understood

In communicating the incentive programs to schools, teachers, and families, a communication protocol was developed and executed through RAMP coaches. Materials (fliers, student logs, teacher logs) were developed, created, and provided for schools. Coaches were trained on the incentive programs and their purposes, and also trained on how to communicate these to teachers and students in selected schools.

The fact that nine RAMP coaches were the agents of communication to approximately 90 schools leaves open the possibility that the communication of the program (despite the assistance of fliers and materials) was not uniform in all schools. Following from this, it is possible that teachers, students, and families understood the goals and the intentions of the incentives program differently, which may have impacted their willingness to participate in the program and their interest in an incentive. Future instantiations of an incentive program may have to reconsider communication efforts in order to ensure that participants uniformly understand the goals of the program and, importantly, how to implement it.

10.5 The use of classroom observation measures

The two treatment arms did not have equally objective measures of treatment effect: Classroom observation measures were employed to determine the effect of treatment on teachers' classroom instructional practices, and the number of books read in a week was used to ascertain whether the incentive encouraged students to read at home. While the use of observational measures was logical, these measures may also have been subject to bias, particularly when compared to a more "objective" measure such as the number of books read in a week by students. The null results (i.e., that the treatment was not associated with increased gains in observational scores from baseline) on the teacher incentive program reported above may, therefore, be partially due to the "noise" that is inherent in observational measures.

11. Conclusions and recommendations

This report presented findings from a pilot study of the effects of a symbolic incentive—a certificate for meeting incentive criteria—on the frequency of student reading at home

and teachers' instructional practices conducted in the Jerash governorate in Jordan. Students in treatment schools were offered an incentive if they were able to read 24 books during the 8-week incentive period (an average of three books per week). Reading books and reading logs were provided for students. Teachers could qualify for an incentive if they received a score of 4 (from a 0–5 scale) on at least half of their classroom observations based on their ability to implement several aspects of the Reading and Math Program.

Overall, students were found to read more frequently in treatment schools as compared to control schools, but this difference was due only to differences in boys' reading frequency. Girl students in treatment schools were not found to read more frequently than their counterparts in control schools. These findings held when we controlled for teacher quality and class differences, but there was no information on student reading prior to treatment.

Nominally, a larger proportion of teachers in treatment schools were found to meet the incentive criteria than in control schools, even when we took into account the pre-treatment differences in teacher practices between treatment and control schools. However, regression models suggested that the treatment was in fact not associated with higher average observation scores (in the linear regression) or an increased likelihood of meeting the incentive criteria (in the logistic regression). A difference-in-difference model confirmed that the growth seen in treatment schools (from 34% to 61% of teachers meeting the criteria) was not statistically different from that seen in control schools (from 24% to 44%) when we controlled for baseline observational scores. In other words, the teacher incentive could not be said to be associated with better implementation of RAMP.

These findings suggest that, despite the limitations presented in this report, symbolic incentives may represent a cost-effective way of motivating certain groups of students to read more frequently outside of school. It should be noted that these effects were seen with little, if any, outside intervention or follow-up with coaches, teachers, and students. It could be hypothesized, therefore, that larger effects might be seen if the program were sustained for a longer duration and if incentives could be given to students more frequently, such as twice per term.

It may also be the case that symbolic incentives would influence students toward other desirable behaviors, or in other contexts. Further research could be conducted to determine if indeed this effect can be observed elsewhere, or with regard to other behavior.

It is less clear what association exists between a symbolic incentive and observed classroom instruction, or teachers' ability to implement prescribed instructional changes. While official symbolic incentives appear to be valued in Jordan, our findings suggest that teachers offered such an incentive for high scores during classroom observation were not more likely than their peers in control schools to receive higher scores. In

addition, the growth seen in implementation scores was not explained by the offer of an incentive. As discussed above, this may have stemmed from the fact that although teachers may have been motivated by official symbolic incentives, effective implementation of RAMP also was due to exposure to that program. Hence, the null findings may be conflating a motivating factor (incentive) with a factor outside of teachers' control (program exposure). Future research on incentives could attempt to untangle these two factors.

Moreover, the observational measures used to determine changes in teachers' instructional practices were "noisy" and generally less reliable than more objective measures (such as those used in the student incentive program). In future this issue could be overcome, in part, by videorecording teachers' lessons and having multiple observers score the teachers, although this approach might significantly increase the cost of the study. Future research also could be used to determine other additional measures—such as surveys, student ratings, or peer ratings—that could produce evidence of teachers' classroom practices and attitudes,.

References

Ahmed, Akhter U., and Carlo del Ninno. 2002. "The Food for Education Program in Bangladesh: An Evaluation of Its Impact on Educational Attainment and Food Security." FCND Discussion Paper 138. Washington, DC: International Food Policy Research Institute.

Alderman, Harold, Daniel O. Gilligan, and Kim Lehrer. 2010. *The Impact of Food for Education Programs on School Participation in Northern Uganda.* Washington, DC: International Food Policy Research Institute.

Allan, Bradley M., and Richard G. Fryer, Jr. 2011. "The Power and Pitfalls of Educational Incentives." The Hamilton Project Discussion Paper 2011-07. Washington DC: Brookings Institution.

Angrist, Joshua D., and Victor Lavy. 2009. "The Effects of High Stakes School Achievement Awards: Evidence from a Randomized Trial." *American Economic Review* 99 (4): 1384–1414.

Arcia, Gustavo. 2014. *Incentives and Accountability in Education: A Literature Review.* Prepared for USAID under the Education Data for Decision Making (EdData II) project, Task Order No. AID-OAA-12-BC-00003 (RTI Task 20, Activity 3). Research Triangle Park, NC: RTI International.

Arcia, Gustavo, Harry Patrinos, Emilio Porta, and Kevin Macdonald. 2011. *School Autonomy and Accountability in Context: Application of Benchmarking Indicators in Selected European Countries.* Washington, DC: Human Development Network, World Bank.

Attanasio, Orazio, Emla Fitzsimmons, and Ana Gomez. 2005. *The Impact of a Conditional Education Subsidy on School Enrolment in Colombia.* London, UK: The Institute for Fiscal Studies, Centre for the Evaluation of Development Policies.

Attanasio, Orazio, Costas Meghir, and Ana Santiago. 2005. *Education Choices in Mexico: Using a Structural Model and a Randomized Experiment to Evaluate PROGRESA.* London, UK: The Institute for Fiscal Studies, Centre for the Evaluation of Development Policies.

Baird, Sarah, Craig McIntosh, and Berk Ozler. 2011. "Cash or Condition? Evidence from a Cash Transfer Experiment." *The Quarterly Journal of Economics* 126: 1709–1753.

Barrera-Osorio, Felipe, Marianne Bertrand, Leigh Linden, and Francisco Perez-Calle. 2008. *Conditional Cash Transfers in Education: Design Features, Peer and Sibling Effects. Evidence from a Randomized Experiment in Colombia.* Washington, DC: World Bank.

Barrera-Osorio, Felipe, and Deon Filmer. 2013. *Incentivizing Schooling for Learning: Evidence on the Impact of Alternative Targeting Approaches.* Washington, DC: World Bank.

Baum, Donald, Laura Lewis, Oni Lusk-Stover, and Harry Patrinos, 2014. "What Matters Most for Engaging the Private Sector in Education: A Framework Paper." SABER Working Paper Series No. 8. Washington DC: World Bank.

Behrman, Jere R., Susan W. Parker, and Petra E. Todd. 2009. "Schooling Impacts of Conditional Cash Transfers on Young Children: Evidence from Mexico." *Economic Development and Cultural Change* 57 (3): 439–477.

Behrman, Jere R., Susan W. Parker, and Petra E. Todd. 2011. “Do Conditional Cash Transfers for Schooling Generate Lasting Benefits? A Five-Year Follow-Up of PROGRESA/Oportunidades.” *Journal of Human Resources* 46 (1): 93–122.

Behrman, Jere R., Piyali Sengupta, and Petra E. Todd. 2000. *The Impact of PROGRESA on Achievement Test Scores in the First Year.* Washington, DC: International Food Policy Research Institute.

Berry, James W. 2013. *Child Control in Education Decisions: An Evaluation of Targeted Incentives to Learn in India.* Ithaca, New York: Cornell University.

Bettinger, Eric P. 2012. “Paying to Learn: The Effect of Financial Incentives on Elementary School Test Scores.” *The Review of Economics and Statistics* 94 (3): 686–698.

Blank, Rolf K., and Nina de las Alas. 2009. *Effects of Teacher Professional Development on Gains in Student Achievement: How Meta Analysis Provides Scientific Evidence Useful to Education Leaders.* Washington, DC: Council of Chief State School Officers.

Blimpo, Moussa P. 2013. *Team Incentives for Education in Developing Countries: A Randomized Field Experiment in Benin.* Stanford, California: Stanford University.

Brixi, Hana, Ellen Lust, and Michael Woolcock. 2015. *Trust, Voice, and Incentives: Learning from Local Success Stories in Service Delivery in the Middle East and North Africa.* Monograph. Washington, DC: World Bank. http://dx.doi.org/10.1596/978-1-4648-0456-4

Brombacher, Aarnout, Jonathan Stern, Lee Nordstrum, Chris Cummiskey, and Amy Mulcahy-Dunn. 2014. *National Early Grade Literacy and Numeracy Survey–Jordan. Intervention Impact Analysis Report. Final Report.* Prepared for USAID under the Education Data for Decision Making (EdData II) project, Task Order No. AID-278-BC-00019 (RTI Task 16). Research Triangle Park, NC: RTI International. http://pdf.usaid.gov/pdf_docs/PA00KH3M.pdf

Bruns, Barbara, Deon Filmer, and Harry Anthony Patrinos. 2011. *Making Schools Work: New Evidence on Accountability Reforms*. Washington, DC: World Bank.

Bruns, Barbara, and Javier Luque, 2014. *Great Teachers: How to Raise Student Learning in Latin America and the Caribbean.* Washington DC: World Bank.

Carnoy, Martin, Iliana Brodziak, Thomas Luschei, Tara Beteille, and Prashant Loyalka. 2009. *Do Countries Paying Teachers Higher Relative Salaries Have Higher Student Mathematics Achievement?* TEDS-M Project Report. Amsterdam: International Association for the Evaluation of Educational Achievement.

Chaudhury, Nazmul, and Dilip Parajuli. 2010. “Conditional Cash Transfers and Female Schooling: The Impact of the Female School Stipend Programme on Public School Enrolments in Punjab, Pakistan.” *Applied Economics* 42: 3565–3583.

Contreras, Dante, and Tomás Rau. 2012. “Tournament Incentives for Teachers: Evidence from a Scaled-Up Intervention in Chile.” *Economic Development and Cultural Change* 61 (1): 219–246.

Demas, Angela, and Gustavo Arcia, 2015. “What Matters Most for School Autonomy and Accountability: A Framework Paper.” SABER Working Paper Series No. 9. Washington DC: World Bank.

De Janvry, Alain, Frederico Finan, and Elisabeth Sadoulet. 2006. *Evaluating Brazil's Bolsa Escola Program: Impact on Schooling and Municipal Roles.* Berkeley, California: University of California, Berkeley.

Duryea, Suzanne, and Andrew Morrison. 2004. *The Effect of Conditional Transfers on School Performance and Child Labor: Evidence from an Ex-Post Impact Evaluation in Costa Rica.* Washington, DC: Inter-American Development Bank.

Filmer, Deon, and Norbert Schady. 2011. "Does More Cash in Conditional Transfer Programs Always Lead to Larger Impacts on School Attendance?" *Journal of Development Economics* 96: 150–157.

Fryer, Roland G., Jr. 2011a. "Financial Incentives and Student Achievement: Evidence from Randomized Trials." *The Quarterly Journal of Economics* 126: 1755–1798.

Fryer, Roland G., Jr. 2011b. *Teacher Incentives and Student Achievement: Evidence from New York City Public Schools.* Cambridge, Massachusetts: Harvard University and the National Bureau of Economic Research.

Fryer, Roland G., Jr., and Richard T. Holden. 2013. *Multitasking, Dynamic Complementarities, and Incentives: A Cautionary Tale.* Cambridge, Massachusetts: Harvard University.

Galiani, Sebastian, and Patrick J. McEwan. 2013. "The Heterogeneous Impact of Conditional Cash Transfers." *Journal of Public Economics* 103: 85–96.

Glewwe, Paul, Nauman Ilias, and Michael Kremer. 2010. "Teacher Incentives." *American Economic Journal: Applied Economics* 2 (3): 205–227. http://www.aeaweb.org/articles.php?doi=10.1257/app.2.3.205

Gorozidis, Georgios, and Athanasios G. Papaioannou, 2014. "Teachers' Motivation to Participate in Training and to Implement Innovations." *Teaching and Teacher Education* 39: 1–11.

Karabenick, S. A., and A. Conley. 2011. *Teacher Motivation for Professional Development. Math and Science Partnership—Motivation Assessment Program.* Ann Arbor, MI: University of Michigan.

Kremer, Michael, Sylvie Moulin, and Robert Namunyu. 2003. "Decentralization: A Cautionary Tale." Poverty Action Lab Paper No. 10. Cambridge, Massachusetts: Massachusetts Institute of Technology (MIT).

Lavy, Victor. 2002. "Evaluating the Effect of Teachers' Group Performance Incentives on Pupil Achievement." *The Journal of Political Economy* 110 (6): 1286–1317.

Leithwood, Kenneth, Rosanne Steinbach, and Doris Jantzi. 2002. "School Leadership and Teachers' Motivation to Implement Accountability Policies." *Educational Administration Quarterly* 38 (1): 94–119.

Levitt, Steven D., John A. List, Susanne Neckermann, and Sally Sadoff. 2013. *The Behavioralist Goes to School: Leveraging Behavioral Economics to Improve Educational Performance.* Chicago, Illinois: University of Chicago.

Levy, Dan, and Jim Ohls. 2007. *Evaluation of Jamaica's PATH Program: Final Report.* Washington, DC: Mathematica Policy Research.

Maluccio, John A., and Rafael Flores. 2005. *Impact Evaluation of a Conditional Cash Transfer Program: The Nicaraguan Red de Protección Social [social safety net].* Washington, DC: International Food Policy Research Institute.

Meng, Xin, and Jim Ryan. 2010. "Does a Food for Education Program Affect School Outcomes? The Bangladesh Case." *Journal of Population Economics* 23: 415–447.

Mourshed, Mona, Chinezi Chijioke, and Michael Barber, 2010. *How the World's Most Improved School Systems Keep Getting Better*. London: McKinsey & Co.

Muralidharan, Karthik. 2012. "Long-Term Effects of Teacher Performance Pay: Experimental Evidence from India." Working paper. San Diego, California: Department of Economics, University of California–San Diego. http://econweb.ucsd.edu/~kamurali/research.html

Muralidharan, Karthik, and Venkatesh Sundararaman. 2011. "Teacher Performance Pay: Experimental Evidence from India." *Journal of Political Economy* 119 (1): 39–77.

Ponce, Juan, and Arjun S. Bedi. 2010. "The Impact of a Cash Transfer Program on Cognitive Achievement: The Bono de Desarrollo Humano of Ecuador." *Economics of Education Review* 29 (1): 116–125. http://dx.doi.org/10.1016/j.econedurev.2009.07.005

Riccio James A., Nadine Dechausay, David Greenberg, Cynthia Miller, Zawadi Rucks, and Nandita Verma. 2010. *Toward Reduced Poverty Across Generations: Early Findings from New York City's Conditional Cash Transfer Program*. New York City: MDRC. http://www.mdrc.org/sites/default/files/full_588.pdf

Schady, Norbert, and Maria Caridad Araujo. 2006. *Cash Transfers, Conditions, School Enrolment, and Child Work: Evidence from a Randomized Experiment in Ecuador.* Washington, DC: World Bank.

Schieb, Laura J., and Stuart A. Karabenick. 2011. *Teacher Motivation and Professional Development: A Guide to Resources.* Math and Science Partnership. Motivation Assessment Program. Ann Arbor, MI: University of Michigan.

Schultz, T. Paul. 2004. "School Subsidies for the Poor: Evaluating the Mexican PROGRESA Poverty Program." *Journal of Development Economics* 74 (1): 199–250.

Uysal, Hacer H. 2012. "Evaluation of an In-service Training Program for Primary-School Language Teachers in Turkey." *Australian Journal of Teacher Education* 37(7). http://dx.doi.org/10.14221/ajte.2012v37n7.4

Vegas, Emiliana, and Ilana Umansky. 2005. *Improving Teaching and Learning Through Effective Incentives: What Can We Learn from Education Reforms in Latin America?* Washington, DC: World Bank.

World Bank, 2010. *Jordan: SABER Teachers Country Report.* Prepared for the Systems Approach for Better Education Results (SABER) program. Washington DC: World Bank.

Annex A: Incentive structures used in selected countries in Latin America

Country and Program	Program Description
Colombia: Estatuto de Profesionalización Docente (Status of Teacher Professional Development [EPD]) 2002	This program encompasses higher standards for new recruits, annual performance evaluations, and promotion based on competency. EPD teachers are better educated and have lower dropout rates, but learning outcomes are mixed. Annual teacher evaluations are weak; directors give all teachers high marks to reduce internal conflict. A more objective approach to teacher evaluation is needed.
Peru: Carrera Pública Magisterial (Public Educator [CPM]) 2008	New recruits take a stringent entry test, and only those scoring 60% and above pass. The CPM entry salary is 50% above the non-CPM salary. CPM teachers have better classroom practices and produce learning outcomes that are significantly higher than those of non-CPM teachers. Teacher evaluations include classroom observation by trained teachers and feedback from directors, peers, students, and parents.
São Paulo State, Brazil: Prova de Promoção (Promotion of Proof) 2009	In this structure, much higher salaries are linked to content mastery. The highest salary is four times the per capita gross domestic product (GDP), which is in the top 10% of professional salaries in the nation. Noted problems include the entry test's lower-than-expected level of difficulty. The system is being reconfigured, and no evaluation of its results has been reported.
Ecuador: Ley de Carrera Docente y Escalafón del Magisterio (Law on Educator Standards and Pay Scales) 2009	This includes mandatory performance evaluations, with dismissal upon refusal. The internal evaluation includes self-evaluation and evaluations by peers and a committee. The external evaluation is based on tests of reading skills, pedagogical skills, and content mastery. Those scoring above 80% receive a salary increase valid for four years. To qualify for this increase again, teachers must be re-evaluated. Those scoring 60–80% receive no salary increase and must be evaluated in two years. Those scoring below 60% must be re-evaluated the following year. If they score below 60% again, they are dismissed. In the first round of exams, of the 2,570 teachers tested, two (0.08%) teachers were rated excellent, 624 (24%) were very good, 1,873 (73%) were good, and 77 (3%) were poor. No evaluation of student outcomes is available.
Mexico: Carrera Magisterial (Teaching Career) 1992	Tests with evaluation thresholds were linked to salary increases of up to 200%. However, test scores represented only 28% of the total points needed to earn this increase. As a result, significant gaming of the system eliminated its impact on performance and student test results.

Source: Bruns and Luque 2014.