DESIGNING EFFECTIVE TEACHER INCENTIVE PROGRAMS
IMPACT EVALUATION UNDER THE BRNED PROGRAM FOR RESULTS

PROPOSAL OBJECTIVES

This study aims to build the evidence on how to effectively incentivize teachers to change behavior in a way that increases student learning outcomes. The study will examine the roles of different approaches to incentives, such as building on extrinsic motivators such as performance rewards or building on teachers’ intrinsic motivation by providing information and feedback on student performance. The study will also (resources permitting) contrast incentives provided to teachers with incentives provided to students.

REVIEW OF EVIDENCE

The rationale for performance-based incentives for teachers rests on the premise that given the structure of their contracts, teachers are poorly incentivized to exert effort. This is because wages are mostly determined by their level of education and experience (Hanushek 1996, Lockheed and Verspoor 1991). Impact evaluation evidence around such programs shows mixed results. Positive impacts on learning have been found in randomized evaluations in developing countries: Kenya (Glewwe et al., 2010) and India (Muralidharan and Sundararaman, 2011), with effects between 0.14 SD and 0.27 SD after two years of implementation and with some spillovers in learning in India. Positive and significant impacts are also seen in a non-experimental study in Israel (Lavy 2002, 2004). Negative impacts of such programs have also been seen in some cases, such as: Portugal (Martins, 2009) and New York City (Fryer, 2011). A recent meta-study of interventions aimed at increasing learning outcomes in primary schools in developing countries concludes that teachers do indeed respond to incentives (McEwan 2013).

The proposed IE aims to shed light on important open questions around the relative effectiveness of teacher incentives for improved school performance.

One critical set of questions relates to the inter-temporal dimension and here there are two key issues: (a) do impacts of teacher incentive programs change when it becomes a repeated game? A-priori impacts of second round incentives could be higher than the first round due to established trust among agents about the credibility of the program or they could be lower due to reduction in the novelty factor; (b) what happens to teacher effort once incentives are withdrawn? Does it remain high or does it diminish, and can it go down to below pre-reward levels due to a reduction in intrinsic motivation in response to the withdrawal of extrinsic rewards? This question is key in light of some recent research which seems to show a limited impact of rewards on “engagement” (current activity) and a negative one on “re-engagement” (persistence) (Benabou and Tirole 2003).

A second set of issues surrounds the question of design – can we expect incentives to be effective if teachers do not really know how to improve performance? To test this, we replicate the design of
Muralidharan and Sundaram (2010) by adding treatment arms that include provision of feedback to teachers on performance in addition to incentives. We extend this research by providing feedback not just on diagnostics around student tests but also diagnostics from detailed classroom observations that use cutting-edge tools in the field (we are using a new and innovative observation tool called TIPPS).

A third set of research questions revolve around testing the relative effectiveness (including cost-effectiveness) of other approaches vis-à-vis teacher incentives. Specifically, we will examine the relative effectiveness of teacher effectiveness against (a) simply providing feedback (on student testing diagnostics and classroom observations) and, budget permitting, (b) providing performance-based incentives directly to students.

The second of these, student incentives, is particularly interesting as it has, to-date, been an under studied dimension of the question. While the body of research that assesses the role of information for accountability and of providing incentives to teachers is growing, evidence on this last dimension – incentivizing students directly – is extremely limited. Only a few studies have rigorously examined the impacts of providing performance-based incentives to students directly in developing countries. One study in Kenya has shown that “merit-based” scholarships where cash transfers were based on successful achievement of performance benchmarks has led to improved test scores (Kremer et al. 2007). Blimpo (2010) finds significant gains in exam scores as a result of individual and team incentives for performance on secondary-school certification exams in Benin. Li et al. (2010) report that a cash incentive linked to performance led to no measurable impacts on learning outcomes in China. Last, a study from Cambodia, has examined impacts of transfers based on baseline performance (and not linked to subsequent effort) has shown positive impacts on learning outcomes (Filmer and Barrera-Osorio 2012). This study will add to this literature by evaluating the effectiveness of low-cost student incentives in Tanzania and more importantly contrasting their effectiveness with teacher incentives.

COUNTRY AND POLICY CONTEXT

In Tanzania, there has recently been a national outcry about declining examination results and poor learning outcomes. In 2012, only 30.7 percent of those who sat for the Primary School Leaving Examination (PSLE) exams passed, down from 58.3 percent in 2011. At the secondary level the pass rate of the Certificate of Secondary Education Examination (CSEE) have also declined in the last few years (see Figure 1). UWEZO (2013) reports that many students in the final grade of primary in 2012 lack basic literacy and numeracy skills sought in the second grade: 47 per cent could not read basic English stories; 26 per cent could not read basic Kiswahili stories and 11 per cent could not perform

---

1 Evidence on student incentives programs in mixed in developed countries. In an evaluation of rewards for performance on high-school matriculation exams in Israel, Angrist and Lavy (2009) find a significant impact on girls, but not on boys. A program rewarding good grades in a Canadian university raised grades for women, but not for men (Angrist et al 2009). Studies yield mixed results in United States: (a) Jackson (2010) finds that combining teacher and student incentives had a significant impact on exam performance; (b) Fryer (2010) evaluates incentives programs in schools in three U.S. cities and finds no evidence that the programs affected student achievement; (c) Bettinger (2010) finds that an incentive program for primary school students in Ohio affected scores in math, but not in other subjects.
simple multiplications. Analysis suggests that the decline cannot be attributed to expansion of access alone. There are critical issues of quality in education service delivery than underpin low and declining learning outcomes and have prompted a national outcry in the country.

**Figure 1: Trends in Enrollment and Exam Results**

![Trends in Enrollment and Exam Results](image)


Against this background, some related education-service delivery issues are as follows:

*Teacher time-on-task is low:* Low rates of teacher time-on-task are major problems in education service delivery in Tanzania. The Service Delivery Indicator survey (SDI, 2011) shows that about one in four teachers is absent from school on any given school day. Even when in school, the teachers are absent from the classroom more than half of the time. On average, students in primary schools in Tanzania are taught two hours a day, and students in urban areas are taught for just one and a half hours a day (instead of the 5 hours and 20 minutes required for primary school teachers).

*Teachers have weak incentives to improve performance:* Findings from an analysis of teacher policies\(^2\) shows that in Tanzania entry requirements, teacher pay, and working condition may not be appealing for talented candidates, signaling teaching as a low-status profession. In addition, this analysis demonstrates clear policy-gaps in terms of teacher motivation. Specifically, professional development of teachers is not systematic and there are no clear rewards linked to good performance. In addition, school leaders have few incentives to improve student achievement. While there are wide disparities within regions and districts, local administrators have no direct incentive or guidance on how to minimize disparities or improve overall achievement.

In February 2013, the President of Tanzania launched the “Big Results Now in Education” (BRNEd) program. The primary rationale for program is to urgently arrest the declining trend in student learning outcomes at primary and secondary levels and reverse it. Program design is based on extensive analytical work and consultations. Over the course of a six week retreat, 34 representatives from government offices, development partners, and stakeholders worked through a participatory process, using the existing body of analytical and policy work, to determine the

---

\(^2\) From the SABER (Systems Approach for Better Education Results) exercise on teachers (2012)
priority interventions that could be implemented within two years or less and would improve education results.

The World Bank is supporting the BRNEd program through their program-for-results instruments (BRNEd PforR, P147486), which is expected to become effective in July 2014.

A key initiative within the BRNEd program (and the PforR) relates to teacher motivation. The BRNEd aims to kick-start a larger teacher perception transformation in Tanzania, starting with recognizing the teachers through non-monetary incentives. At Government’s request, part of World Bank support will be geared towards a rigorous impact evaluation of the teacher incentive program, to inform future program and policy design in this area.

**DESCRIPTION OF THE INTERVENTION**

The interventions focus on improving student learning outcomes through performance-based incentives in Tanzanian secondary schools. They are a part of a Government-led pilot that is being intensively supported by World Bank and is expected to help define the design of a teacher incentive program implemented through the BRNEd program.

The study is divided into two phases. In Phase I, a performance-based incentive scheme for teachers was implemented. A total of 420 schools (public and private) in three regions were divided into three groups – one group received performance-based teacher incentives, a second group received both teacher and student incentives, and the third group served as control.

Teacher incentive schools were divided into homogeneous groups based on geographic sub-area and performance at baseline (20 schools per group). Within each group, the top three schools with the highest index score (which combines the average score with the average gain in scores relative to baseline across all three subjects) are slated to receive widely publicized recognition awards. In addition, subject specific recognition awards will be given to the best three teachers in a particular subject within the homogenous group. In student incentive schools, top students also receive rewards. These interventions are at the stage of endline testing. Actual awards were distributed in Feb 2014.

The Phase II of these interventions has been designed in response to specific requests from the Ministry of Education as well as research questions that have arisen during implementation of phase I. There is substantial interest in measuring longer term impacts of these incentives and to this end Government plans to implement one more year of similar teacher awards in a manner that allows us to examine questions around inter-temporal dynamics of such awards. This second year implementation will also be used as an opportunity to identify optimal design choices by experimenting with provision of performance-related feedback to teachers and directly incentivizing students (without confounding those incentives with those provided to teachers).
The proposed evaluation aims to address the following overarching question: how to design performance-based incentive schemes to maximize impacts on learning outcomes? Within this broader research question, the following specific questions will be explored:

**Inter-temporal dynamics of teacher incentives**

1. Is second round impact of teacher incentives bigger (due to established trust around program) than first round impact?
2. Do extrinsic awards negatively impact intrinsic motivation (does teacher effort go down to below pre-award levels once awards are withdrawn)?

**Improving design of teacher incentives**

1. Is the impact of teacher incentives enhanced when teachers are provided with feedback on how to improve performance?

**Contrasting effectiveness of other interventions with teacher incentives**

1. Relative effectiveness of providing teachers with feedback vs incentives.
2. Relative effectiveness of providing incentives to students vs teachers (or in combination).

**Outcome Indicators**

- Student learning outcomes
- Teacher attendance
- Student attendance
- Teachers’ classroom behavior, collected via classroom observation
- Student and teacher perceptions

**Evaluation Design**

Evaluation design is experimental and uses a Randomized Control Trial approach. The evaluation will be conducted 420 schools (public and private) from three regions of the country: Lindi, Morogoro, and Shinyanga. The evaluation has been ongoing for two years (Phase I) and data from first follow-up will be available by Jan 2014. The proposal is to extend the evaluation over two more years (Phase II) to study longer term impacts of the interventions and identify optimal design choices.

Design of Phase I and Phase II are illustrated in Figure 2:
Inter-temporal dynamics of teacher incentives

1. Is second round impact of teacher incentives bigger (due to established trust around program) than first round impact? B vs. H
2. Do extrinsic awards negatively impact intrinsic motivation (does teacher effort go down to below pre-award levels once awards are withdrawn?) G vs. H

Improving design of teacher incentives

1. Is the impact of teacher incentives enhanced when teachers are provided with feedback on how to improve performance? H vs I

Contrasting effectiveness of other interventions with teacher incentives

1. Relative effectiveness of providing teachers with feedback vs incentives. B vs F²
2. Relative effectiveness of providing incentives to students vs teachers (and in combination). B vs E (and C)

---

3 We presume that the external education environment does not change between Phase 1 and 2. Given that the time lag between the two phases is exactly one year, this seems a reasonable assumption

4 Ibid
DATA COLLECTION

Under the IE, the following types of data will be collected:

- Student assessments using curriculum-based standardized in Math, English, and Kswahili
- Head-teacher interviews
- Teacher interviews, perception surveys, and psycho-social modules
- Classroom observations (using cutting-edge new modules) through unannounced visits
- Teacher and student attendance through unannounced visits
- Teacher content knowledge test (to be confirmed)
- Community-perception data (to be confirmed)
- Quantitative and qualitative data on school management committee meetings, community-meetings/discussions/actions

Intermediate outcome indicators:

Teachers’ classroom behavior, collected via classroom observation; teacher satisfaction, collected via teachers’ questioners.

Parental perception about quality of education service delivery, community/parent-participation in school management, quality of school management

Final outcome indicators:

Student and teachers attendance collected both by unannounced visits to schools; student academic achievement collected via students tests.

TEAM

- Shwetlena Sabarwal (Economist, World Bank)
- Deon Filmer (Lead Economist, World Bank)
- James Habyarimana (Georgetown University)
ANNEX: ADDITIONAL INFORMATION

RISKS TO EVALUATION DESIGN

The following measures will be taken to reduce the risk of selection bias, contamination and attrition:

- All schools in the treatment group will be provided with the corresponding treatment. The MoEVT has given the IE team permission to plan and supervise the delivery of school report cards and teacher awards.
- Counterparts from MoEVT were involved in the randomization of schools and have received training on impact evaluation (the IE team from MoEVT participated in a SIEF organized impact evaluation workshop in Addis Ababa in May 2012).
- Schools in the sample are not very close to each other. Therefore risks of contamination are low.
- The risk of attrition for these interventions is negligible – given that implementation of the treatments is under the control of the IE team. In addition, there is minimal turnover of teacher and student cohorts over the two year duration of the study. However, in the event that students or teachers attrite, we intend to weight observations in the analysis with the inverse probability of attrition (as in Fitzgerald, Gottschalk and Moffit 1998)

SAMPLING

Sampling of schools was undertaken using Tanzania EMIS data from 2011 in collaboration with MoEVT. The following steps were undertaken:

- Based on consultations with the MoEVT, three regions were selected for the IE – Shinyanga, Lindi, and Morogoro.
- A complete list of schools (public and private) in these regions was drawn up using EMIS data. These data were validated/buttressed using school mapping data (Dec 2012) undertaken specifically for this IE.
- Schools in each region were divided into homogeneous groups based on geographic sub-area and performance at baseline. This is to ensure that all teachers have a decent shot at winning the awards.
- Schools within each homogeneous group were randomly divided into 3 treatment groups and 1 control group. This was done as follows:
  - Construct a natural category (region, district, or combination of districts) with about 60 schools.
  - Within, natural category: generate a ranking of school quality using pass rates; Create three tournaments (terciles of the ranking).
- Within each tercile: randomly sample half of schools to receive teacher incentives - other half becomes a 'Teacher incentives control'
- Stratify each tercile by assignment to teacher incentives
- Within each of the two tercile strata: randomly assign half of schools to school report card (SRC) and the other half to pure control