SOCIAL COMPARISON, REWARDS AND INCENTIVES TO LEARN: A RANDOMIZED CONTROL TRIAL IN UGANDA*

DAGMARA KATRENIAKOVA§

CERGE-EI¶

Abstract

Substantial progress has been made in improving access to schooling in developing countries. Nevertheless, higher enrollment needs to be accompanied by advances in education quality in order to avoid stagnation or, at worst, education quality downturn. Large number of interventions has been implemented with the aim to lower absenteeism and improve students' performance. Recent attention has been paid to enhancing students' motivation, using both financial and non-financial rewards but little has been said about symbolic rewards. This paper contributes to the discussion threefold. First, it studies the effect of symbolic rewards, such as comparative feedback revealing students' group performance given in a form of a report card. Students repeatedly facing such comparative feedback improved their performance within an academic year. Second, it allows for the direct comparison of two types of social comparative feedbacks - within and across class group comparisons. The results suggest no significant difference in type of comparison provided. Last, it helps to understand the value added of financial and reputation rewards introduced into social comparison framework. The effects of social comparison treatment become more pronounced once the real rewards are introduced. While financial rewards seem to motivate all students with or without social comparison treatment to perform better, reputational rewards have no effect on students' performance.

Keywords: education, motivation, financial rewards, reputational rewards, randomized control trial, competition, group evaluation, Uganda
JEL Classification: C90, C93, D04, I21, I29, O55

---

* All errors are mine. The research was supported by GA UK grant No. 338911. Dagmara Katreniakova further acknowledges support by GDN grant No.60.
§ Email: dkatreni@cerge-ei.cz
¶ CERGE-EI, a joint workplace of the Center for Economic Research and Graduate Education, Charles University in Prague, and the Economic Institute of the Academy of Sciences of the Czech Republic, v.v.i.

1. Introduction

Higher school enrollment, the second Millenium Development Goal supporting primary education for all children in developing countries, needs to be accompanied with advances in education quality in order to achieve sustainable improvement. Among the different approaches to education quality improvement, recent attention has been paid to information provision and social comparison. According to social comparison theory, informing a child about his/her performance without comparing it to other children causes unstable evaluations of the child's ability and can influence effort negatively. On contrary, comparison enables a child to find his/her relative position within a particular group which can lead via enhanced competitiveness to increase in effort and performance improvement. In this project I study whether the provision of comparative feedback, a pure information without any further incentivization, can increase students effort and lead to performance improvement. Moreover, different types of comparisons can motivate students differently and thereby result in diverse changes in students' performance. In treatment group 1, I introduce intra-class competition based on the comparison of small group of students within their class. In treatment group 2, I introduce inter-class competition based on the comparison of different classes between each other. The design of the experiment allows for direct comparison of the two treatment groups. Furthermore, I analyze to what extent the results differ if students are rewarded financially versus non-financially. The rewards were orthogonally introduced to existing social comparison framework before the last testing round (five testing rounds were conducted in total). Such design helps to differentiate the value added of the rewards. Moreover, since students were evaluated in groups, peer effects can influence also dropout and absence rates, which I analyze as secondary outcomes.

2. Related literature

The quality of education is mostly measured using student test scores, students' and teachers' absence rates, enrollment, survival and drop-out rates. Absences are measured mostly during unexpected visits at schools (Chaudhury el al. (2006), Kremer et al. (2004), Alcazar et al. (2006)) and have been found having negative impact on students' performance. Miller, Murnane and Willey (2008) for example have discovered that ten-day teachers' absence lowers students' achievement in Math by approximately 3.3 per cent on average. Naturally the question of interest has been how to lower absenteeism at schools. Vermeersh (2003), Linden, Shastry (2005) or Banerjee, Barnerji, Duflo, Glennerster, Khemani (2008) and others focus primarily on improvement studying conditions at schools. Interventions used in these studies comprise subsidized meals and other nutrition programs, training community members in a new testing tool together with running additional classes of reading by volunteers respectively. While some improvements have positive impact on attendances (e.g. remedial reading classes, provision of food), the other leave scores unchanged (e.g. training of local community members). Existing evidence also suggests that monitoring can help to lower the absence rates but the success depends on the type of monitoring. As Banerjee, Duflo (2005) conclude, while impersonal monitoring seems to work (e.g., Duflo and Hanna (2006)) the opposite holds for personal monitoring (e.g., Kremer and Chen (2001)). Within class and across class comparison of students evaluated as groups, according to the project results, seems to influence students' absence and dropout rates. Peer pressure and responsibility towards other group members seem to be the driving force of the results. The fact that peers can enhance students' involvement is shown by Mas and Moretti (2009) and Falk and Ichino (2006). The drawback of the design is that both dropout and absence rates are based on our school visits only due to time and financial constraints.

The fact that even if teachers are present they do not necessarily teach (Kremer et al. (2004)) has inspired other authors to focus on the question how to improve students' performance instead of students or teachers presence. Similarly, first the authors looked at the improvements in the input side. Glewwe, Kremer, Moulin (2002), Glewwe, Kremer, Moulin, Zitzewitz (2004), Banerjee, Cole, Duflo, Linden (2003) or Vermeersch (2003) studied the impact of improving school conditions on students' performance. In their studies the treatment was based on additional textbooks for schools, equipping classrooms with flipcharts, remedial courses together with computer assisted learning program or provision of subsidized meals respectively. The results differed[1] and led researchers to focus more on students' and teachers' motivation.

The most commonly used incentive tools are financial rewards (directly cash money or financial vouchers) for both, teachers and students (Angrist et al (2002, 2006), Kremer, Miguel, Thornton (2002), Bettinger, Kremer, Saavedra (2010), etc). They all find that students on average react positively on such monetary rewards and on average work harder which suggests that students can improve their performance if they are motivated. However, once these results are decomposed into groups based on the level of children performance, usually low and high performing students do not improve and only middle-range-performance students, who with some additional effort can compete with high performance students, increase their scores. In order to try to motivate all students, different rewarding scheme has been presented by Blimpo (2010). He designs an experiment when he evaluates and financially rewards students based on group performance and compare them to performance of individuals under the same treatment. He finds positive effect of group evaluation on all-range-performance students, especially if they are included in a tournament.

Further interest lies in studying the effects of non-pecuniary incentives, especially status or social recognition rewards; however the number of such empirical studies is limited. Kosfeld and

[1] While some of these improvements result in higher students' test scores (remedial courses, subsidized meals), the other leave scores unchanged (flipcharts, additional textbooks, computer assisted learning program).

Neckerman (2011) designed a field experiment where students in the treatment group are offered symbolic rewards (congratulatory card) for their work performance while students in control group are not offered anything. Their results provide strong evidence that status and social recognition rewards have motivational power and lead to increase in work performance. Information in a form of report card (feedback) could serve as a symbolic status reward, too. Andrabi, Das and Ijaz-Khwaja (2009), for example, provided parents, teachers and headmasters with report cards informing them how children are doing in a particular school. The intervention resulted in 0.1 standard deviation improvement in students' test scores. On contrary, Erickson et al. (2009) did not find any effect of feedback of any intensity to influence employee's performance. Bandiera (2011) finds negative effects of feedback provision. Workers in his experiment lower their performance and increase their dropout from work after being exposed to feedback, which informed them about how they ranked in terms of their work performance compared to their colleagues. The results of feedback on performance are diverse and according to my opinion depend crucially on the local circumstances.

The remainder of the paper is organized as follows. Section 2 summarizes the methodology; it describes randomized control trials and randomization scheme, experimental design, logistics of the experiment, overall sample as well as data summary. Section 3 summarizes the results of the intervention. Section 4 concludes the main findings.

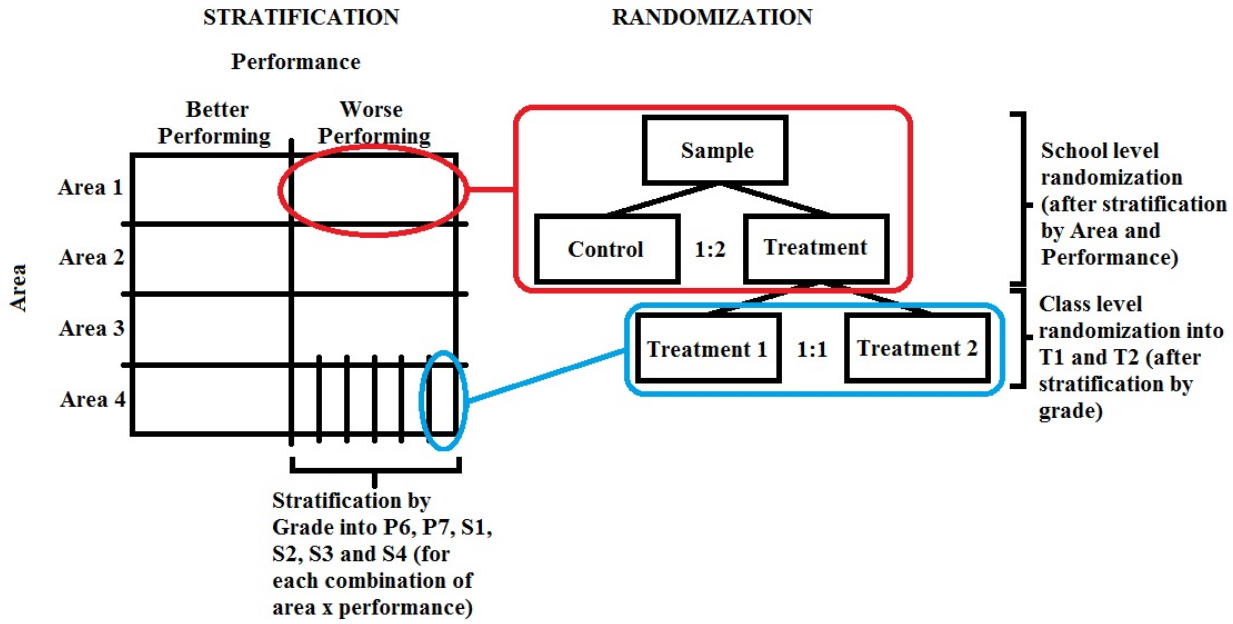
3. Methodology

3a. Randomized Control Trials and randomization

To evaluate the effect of the intervention, I designed Randomized Control Trial (RCT) experiment. If the randomization is done properly with high enough numbers of random draws, it ensures balance between a control and a treatment group in expectations in terms of observables as well as unobservables, and therefore there should be neither selection bias nor confounding

factors to spoil the program evaluation. In order to increase balance between control and treatment groups, the sample can be stratified along crucial dimensions. In my case, I stratified the sample along three dimensions - the location of the school (especially distance from Kampala, Ugandan capital city), students' level (grade 6 and 7 of primary education and grades 1 up to 4 of secondary education) and the schools' results in the national level examination[2]. Every year students of P7 in primary schools and S4 in secondary schools take the leaving examinations that are compulsory in order to complete their study and to proceed to higher level. Using the data on PLE and UCE, I was able to divide schools into better and worse performing schools. Stratification divided my sample into 48 stratas. Within each strata, I randomized the sample into treatment and control groups.

Figure 1: Stratification and randomization scheme



[2] Uganda introduced Universal Primary Education (UPE) for all in 1997, allowing up to four students to go to school for free. Later it was extended to all children. Primary education is a seven-year program and for successful completion students need to pass the national Primary Leaving Exam (PLE) at the end of grade 7. Without passing PLE they cannot be admitted to a secondary school. Secondary school consists of two levels - "O-level", which is four year program from S1 up to S4 completed by passing Ugandan Certificate of Education (UCE); and "A-level", which is a two year extension to the O-level and is completed by passing Ugandan Advanced Certificate of Education (UACE). In 2007 Uganda introduced Universal Secondary Education (USE) as the first African country. The school year consists of 3 trimesters and lasts from January until December. Students are supposed to be examined by midterm and final, however, students do not necessarily have access to their evaluations and have limited information about their improvements.

The randomization was done in two stages (as shown in Figure 1). First, after stratification of the sample by school performance and area, I randomized the whole sample of 53 schools into treatment and control group in a ratio 2:1. The randomization was done at the school level and resulted in 36 treatment schools and 17 control schools. School level randomization in the first stage was chosen in order to minimize control group contamination due to information spillovers, which could happen in case there were both treatment and control groups within one school. In the second stage, after stratification the overall treatment group by school level, I divided it randomly into treatment 1 (T1) and treatment 2 (T2) in a ratio 1:1. In this scenario, when a school belongs to a control group, none of its students receive any treatment. However, if the school belongs to a treatment group, then its classes can receive any combination of T1 and T2. Another possibility was to randomize purely at the school level. The advantage would be that all the classes within one school would receive the same treatment. However, it was not feasible in my case due to budget, the sample size but also time constraint.

Overall, 1/3 of the sample is the control group, 1/3 is treatment group 1 and 1/3 is treatment group 2. Exposure to the treatment is the only difference in the outcomes between the control and treatment groups.

3b. Experimental design

Two types of social comparisons were introduced - intra-class (or within-class) comparison (treatment 1, T1) and inter-class comparison (treatment 2, T2). Students in treatment 1 were randomly divided into groups of three to four classmates within each class and were evaluated as groups within their respective classes. It means, group averages were taken into account when comparing the students' performance. Students in treatment 2 were evaluated as a whole class (using class average) in comparison to other classes of the same grade in different schools.

Students were tested in Math and English every one and a half months. In order to ensure transparency, I used own constructed tests. After the evaluation of the exams, students in treatment groups were given the feedback. While students in T1 received the results in the subsequent visit, students in T2 received their first results with one visit delay (due to the number of exams to be evaluated). The feedback was provided to students in the form of a report card, which was sticked onto a small progress report book each child in the treatment group received from us. The books contained all necessary information to keep a child's attention and motivation active. The content of the report card was piloted during the baseline survey and designed in a way that all students in primary and secondary schools should understand it. In addition to that, each team member made sure while disseminating the report cards that students understood the feedback. The books were stored at schools and I was ensured by the school management that students had free access to them. At the end of the academic year, children kept their books.

Feedback differed across treatment groups with respect to its content. Each student in treatment 1 received information about how he scored in Math and in English together with the results of two to three classmates from his group and the position of the group within his class. Furthermore, started from testing round 3, the student received information about how he (and his group members) improved or worsened in between two preceding testing rounds. Students in treatment 2 received information about how they scored in Math and in English personally (i.e., how much they contributed to the class average) and the position of their class compared to other classes in the sample. Students in control group did not receive any information. Students were not offered further rewards until testing round 4 was finished.

Once the last student was tested in testing round 4, I re-randomized the sample orthogonally into financial/reputational/no-reward groups. The randomization was done at school level in order to avoid spillover effects and possible confusion. Therefore, all classes within one school received one type of rewards only. The aim was to observe whether introduction of additional rewards could

enhance students' performance, especially if interacted with the treatments T1 and T2. In order to announce the competition, I organized additional meetings with students to explain the conditions in details. Moreover, I left fliers in their classrooms so that their absent classmates learn about the competition, too. Students in financial treatment could win 2000UGX per person (which is approximately 0.80 US cents according to current exchange rate). Students in reputational reward scheme were promised that if they qualify for the reward their names would be announced in local newspapers Bukedde (the most popular in the region). The qualification criteria differed based on original randomization into treatments (see Table 1) but the general rule was to reward top performing students/groups/classes as well as the most improving students/groups/classes.

Table 1: Qualification criteria for winning the rewards

	Financial rewards (2000 UGX)	**Reputational Rewards (Winners' names published in local newspapers)**	**No rewards**
Within-class social comparison (Treatment 1)	15% of best performing and 15% of best improving **groups**	15% of best performing and 15% of best improving **groups**	Pure within-class social comparison group, no rewards
Across-class social comparison (Treatment 2)	15% of best performing and 15% of best improving **classes**	15% of best performing and 15% of best improving **classes**	Pure across-class comparison group, no rewards
Control group	15% of best performing and 15% of best improving **students**	15% of best performing and 15% of best improving **students**	Pure Control Group, no rewards

Note: In order to avoid confusion, students were given exact information regarding the number of winning groups (if in T1), the number of winning classes (if T2) and the number of winning students (if originally in control group). I used percentages in order to guarantee comparable number of winners across all treatment groups.

Students were tested twice per term. Testing date and time were arranged 10-14 days in advance with the headmaster or the director of the school and confirmed a day before the testing. In general, three to four schools were visited per day 5 times a week. The research team consisted of four team members and each of them visited one class per school. The agenda of each visit was

similar. After we entered the class, students were given "Before Math questionnaire"[3], then they had 30 minutes to finish Math exam. Afterwards students answered "After Math Before English questionnaire"[4], filled English exam in the next 20 minutes and at the end they filled "After English questionnaire"[5]. In treatment schools, students were given report cards at the very beginning before we started with questionnaires and examinations. Team members were trained to explain the contents of the report cards such that all students understood their message. Teachers were allowed to be present in the class. Their attendance was helpful especially at the beginning during the dissemination of the report cards as the students had a tendency to cheat and to present themselves under different names. They also maintained discipline in the class. Nevertheless, they were kindly asked not to intervene into the testing at all.

3c. Timing and Final sample

The experiment took two years. Baseline survey was conducted between September and December 2011. The intervention implementation and the core data collection took place from January 2012 until December 2012. Follow-up session was arranged between May and August 2013.

The main task of the baseline survey was to explain the project to headmasters, to agree on cooperation and to interview students. In total we visited 60 schools from three districts (Wakiso, Mukono and Buikwe) and interviewed 8158 students from seven different grades (P5 to P7 in primary schools and S1 to S4 in secondary schools). Students were asked questions regarding their sex, age, parental background (job, education), family background (family decomposition and

[3] [4] [5] The core questions of the questionnaires were students' expectations regarding how many points they thought they would obtain from Math and English examinations, how much effort they planned to put/they put into answering the questions and the level of their current happiness. All of these questions we asked them before as well as after each exam.

family wealth), students' education, health, interests, self-esteem and students' aspirations. While the students in primary schools were given a questionnaire in Luganda, secondary schools students were asked questions in English. After we entered the classroom, each child was given a pen to make sure that they can fill in the questionnaires. At least one team member was available in every classroom during the questionnaire. Each team member was obliged to answer and clarify any questions students raised. Children were rewarded with a sweet after they finished.

From January 2012, I limited the sample to two districts only – Buikwe and Mukono. (Based on baseline survey, schools from Wakiso district (the district closest to Kampala) were too few and too different from the rest of the sample). All schools in the sample were connected to local non-governmental organization called Uganda Czech Development Trust (UCDT). UCDT is a local affiliation of the non-governmental organization Archdiocese Caritas Prague, Czech Republic, which has been running a sponsorship program "Adopce na dalku" in Uganda since 1993. Students are located into primary and secondary schools based on their own choice, therefore supported students should not differ from not supported students in terms of their school choice.

The final sample consists of 53 schools, 31 primary and 22 secondary schools out of which 19 are public, 24 are private and 10 are community schools. All schools describe their location as rural. The sample comprises 150 classes summing up to 7131 students (as of the testing round 1) from six grades (P6 and P7 in primary schools, S1 up to S4 in secondary schools).

Table 2: Project's timeline

					Reward scheme introduced ↓			
2011 Baseline Survey		**2012**						**2013 Follow-up Session**
		Testing 1	**Testing 2**	**Testing 3**	**Testing 4**	**Testing 5**		
Students, teachers and headmasters interviewed	BREAK	Baseline testing from Math and English and question-naires; No treatment	**T1** received first treatment; **T2** no treatment	**T1** received treatment including improvement status; **T2** received first treatment	**T1** received treatment including improvement status; **T2** received treatment including improvement status	**T1** received treatment including improvement status; **T2** received treatment including improvement status; Chosen students competed to win prizes	BREAK	No treatment provided, students examined from Math and English;
						↑ Rewards disseminated		

Note: T1 (treatment 1) stands for within-class social comparison treatment; T2 (treatment 2) represents across-class social comparison group; Reward scheme was introduced after testing round 4, students in selected schools were informed that they will be rewarded (with financial or reputational reward) if they qualify for the prize. Qualification criteria differed based on initial randomization (T1,T2,C).

4. Results

4a. Randomization Balance

Properly done randomization ensures that treatment and control groups are in expectations the same, in terms of both observable and unobservable characteristics. In the following tables I show (im)balance in baseline characteristics collected before any treatment was implemented. In my scenario, due to the randomization at school level, I "tossed a coin" 53 times, which may or may not be considered as high enough.

The core variables of interest are students' results from Mathematics and English exams. Tables 3a and 3b show that neither of the differences in scores from Mathematics, English and their sum is

statistically different between treatment groups T1 and T2 and the control group as well as between rewarded and non-rewarded groups.

Table 3a: DIFFERENCES IN MEANS BETWEEN SOCIAL COMPARISON TREATMENT AND CONTROL GROUPS

Variable	T1 mean	T2 mean	C mean	(T1-C) std.err.	(T2-C) std.err.
Mathematics	11.015	11.198	11.092	-0.077	0.106
				(0.99)	(0.96)
English	11.551	11.927	11.477	0.074	0.450
				(1.53)	(1.72)
Sum of Mathematics and English	22.566	23.125	22.569	-0.003	0.556
				(2.30)	(2.43)

T1 stands for within-class comparison, T2 across-class comparison, C control group. Robust standard errors adjusted for clustering at school level are in parentheses. * significant at 10%; ** significant at 5%; *** significant at 1%

Table 4b: DIFFERENCES IN MEANS BETWEEN REWARD TREATMENT AND CONTROL GROUPS

Variable	FinRew mean	RepRew mean	No Rewards	(Fin – No) std.err.	(Rep – No) std.err.
Mathematics	10.038	11.200	10.231	-0.193	0.969
				(0.94)	(0.88)
English	11.039	11.215	10.151	0.889	1.064
				(1.75)	(2.11)
Sum of Mathematics and English	21.077	22.416	20.382	0.696	2.034
				(2.27)	(2.69)

FinRew stands for financially rewarded group, RepRew for reputationally rewarded group and No Rewards represents control group with no rewards. Robust standard errors adjusted for clustering at school level are in parentheses. * significant at 10%; ** significant at 5%; *** significant at 1%

The baseline questionnaires together with data on participation offer further tests for testing treatment-control imbalances. The results shown in Tables 4 show the differences in response rate to baseline questionnaires between treatment and control groups. None of the differences is significant. To see to what extent students differ in terms of their results from Mathematics and English together, when answering baseline survey questions, see Appendix B. The table shows whether similar students in terms of their performance answered survey questions similarly. The Table 4 contains also testing for the difference in the participation during testing round 1 and

revealed that the number of participants did not differ significantly between treatment and control groups. Since the number of participants can serve as a proxy for class size, one can assume that randomization resulted in comparable class sizes across treatment and control as of baseline data. Randomization therefore successfully divided the sample and it can be assumed that treatment groups are on average the same in terms of observables as well as unobservables as control group.

Table 5: DIFFERENCES IN BASELINE VARIABLES BETWEEN TREATMENT AND CONTOL GROUPS
(For full scale comparisons see Appendix)

	After Math Questionnaire		After English Questionnaire	
Variable	**(T1 – C)**	**(T2 – C)**	**(T1 – C)**	**(T2 – C)**
Question 1: Expected number of points	-0.166	-0.020	-0.070	-0.010
	(0.21)	(0.09)	(0.29)	(0.16)
Question 2: Subjective effort level	-0.024	0.007	0.017	0.035
	(0.08)	(0.03)	(0.06)	(0.03)
Question 3: Perceived difficulty	-0.051	0.034	-0.027	-0.013
	(0.08)	(0.04)	(0.08)	(0.04)
Question 4: Subjective level of happiness	0.287	0.132	0.274	0.114
	(0.30)	(0.14)	(0.19)	(0.09)
Other variables available in Round 1	-4.335	-0.683		
Participation in round 1	(5.3)	(3.18)		

Note: Robust standard errors adjusted for clustering at school level are in parentheses. The table shows the difference in response rate between treatment and control group. * significant at 10%; ** significant at 5%; *** significant at 1%

Common features of students in developing countries (especially in rural areas) are high drop-out and absence rates and it is not an exception in my data. The attrition (absences and dropouts together) differs between treatment and control groups (non-random attrition) suggesting that the intervention influenced attrition. The table shows that the social comparison treatment (within/across class competition) led to 2-3.7% decrease in dropout rates and 5-6.8% decrease in absence rates, which means the attrition in treatment groups is lower by 7-10% on average compared to the control group.

Table 6: DIFFERENCES IN DROPOUTS/ABSENCES BETWEEN TREATMENT AND CONTROL GROUPS

Variable	(T1 – C) std.err.	(T2 – C) std.err.
Dropout rate	**-0.037****	**-0.020****
	(0.02)	(0.01)
Absence rate	**-0.068***	**-0.050*****
	(0.04)	(0.02)
Attrition	**-0.105****	**-0.070*****
	(0.04)	(0.02)

Dropout is defined as any combination of two to three participations in first three testing rounds but no participation in the last two testing rounds. One single visit is considered as a speculative behavior and is excluded. Student was absent if she missed the last testing round but not first one (or vice versa). T1 stands for within-class comparison, T2 for across-class comparison and C for control group. Robust standard errors adjusted for clustering at school level are in parentheses. * significant at 10%; ** significant at 5%; *** significant at 1%

4b. Probability to dropout and to be absent

During the academic year, I visited schools five times. The participation rates in each testing round differed. There are several reasons. Some students did not have money to pay the school fees and they decided to change the schools to avoid repaying their debt, others changed the school for other reasons (family moved to different area, they were sent to live with other family members, etc.), some completely dropped out of school, some just subscribed as new students and some of the students died. Due to the constraints of the experiment, all participation data are based on our visits only (it means, no random checks of attendances were organized).

In order to distinguish absenteeism from dropouts, I defined the following measures. Students who were present twice or three times in our testing but all their participation happened during first three visits and they did not participate in the last two visits belong to dropout group. Students who participated only once during the entire year (mostly because they did not mention their names correctly or they used nicknames) belong to speculative group. And students who either participated in the first testing but not the last one (with any combination of participation during testing rounds 2, 3 and 4) or vice versa belong to absence group.

Imbalance between treatment and control groups in terms of absence rates, dropout rates and overall attrition rates indicates that the treatment might have an influence on students' participation at school. To measure the impact, I estimated students' probabilities to dropout, to be absent and overall probability not to participate (which I call generally as attrition). The results are presented in Tables 7 and 8. While in the Table 7 I assume fully orthogonal treatments, in the Table 8 allow for treatment interactions. For each measure there are two columns – one with area 1 included and the other one with area 1 excluded[6]. The reason is that the randomization into reward treatments was done just within areas 2, 3 and 4 due to logistic reasons.

Table 7: OBSERVING THE EFFECTS OF TREATMENTS ON PROBABILITY TO DROPOUT, ABSENCE AND ATTRITION RATES

Dependent variable: probability of dropout/absent/attrite	**Probability to Dropout area 1 incl.**	**Probability to Dropout area 1 excl.**	**Probability of Absence area 1 incl.**	**Probability of Absence area 1 excl.**	**Probability of Attrition area 1 incl.**	**Probability of Attrition area 1 excl.**
Within class social comparison (treatment 1)	**-0.041*** ** (0.01)	**-0.043** ** (0.02)	-0.016 (0.02)	-0.009 (0,03)	**-0.064** ** (0.03)	-0.060 (0.04)
Across class social comparison (treatment 2)	**-0.037*** ** (0.01)	**-0.041*** ** (0.01)	-0.018 (0.02)	-0.033 (0.03)	**-0.061** ** (0.03)	**-0.081** ** (0.04)
Financial Rewards	-0.027 (0.02)	-0.030 (0.02)	-0.031 (0.03)	-0.034 (0.03)	-0.066 (0.05)	-0.069 (0.05)
Reputational Rewards	**-0.033** ** (0.01)	**-0.036** ** (0.02)	-0.033 (0.03)	-0.036 (0.03)	**-0.078*** (0.05)	**-0.082*** (0.05)
Baseline Mathematics and English score	**-0.021*** ** (0.00)	**-0.025*** ** (0.01)	**-0.047*** ** (0.01)	**-0.038*** ** (0.01)	**-0.073*** ** (0.01)	**-0.069*** ** (0.01)
X (other variables)	…	…	…	…	…	…
N	7108	5042	7108	5042	7108	5042

Note: Robust standard errors adjusted for clustering at school level are in parentheses. Controlled for stratum (performance, grade and area) fixed effects. X stands for set of stratification variables controlled for in the regression – area (four different areas), school performance at national examination and grade level (P6,P7, S1 up to S4). N stands for the number of observations. Full table can be found in the Appendix.
* significant at 10%; ** significant at 5%; *** significant at 1%

[6] Area 1 is located in Mukono district and it is the area closest to Kampala. Due to time and financial constraints I was not able to reach the area and introduce financial and reputational reward scheme. Therefore, the randomization into reward scheme was done in all areas but area 1.

Both within class competition (T1) and across class competition (T2) helped to lower the dropout rates by approximately 4 %. Surprisingly, while financial rewards did not influence on average the probability to dropout, reputational rewards lowered the probability by 3.4%. None of the treatment effects influenced students' absence rates, suggesting that the reason for their absences were of more priority. The effect on overall attrition is therefore driven by lower dropout rates.

The effect on dropout rates varies more once the interaction of the treatments is allowed. Within class competition without further incentivization decreases the probability of dropout by 2.5%. Adding additional rewards (both, financial and reputational), however, enhances the effects. While additional financial rewards reduce the dropout rates by 4.8%, the reputational rewards reduces the dropout rate by 5.7%. Across class competition itself (without further rewards) decreased the probability to dropout by 2.8% but adding financial rewards does not seem to have any additional impact to lower the dropout rate, on contrary, it seems moderate the effect of pure T2 treatment. Reputational rewards, however, significantly contribute to lower the dropout rates on top of the across class competition (T2).

Absence rates do not seem to be influenced by the interaction of treatments either. Among all eight scenarios it was only reputational rewards combined with across class competition that significantly decreased the probability to be absent. All other scenarios seem to have on effect.

Overall, according to the results it seems that while the reputational rewards motivate students to stay at school more often, especially if combined with either within or across class competition, the effect of financial rewards is not that straightforward as it seems to work only when one let small groups compete within class (T1).

Table 8: THE EFFECTS OF TREATMENTS ON PROBABILITY TO DROPOUT, ABSENCE AND ATTRITION RATES

Dependent variable: probability of dropout/absent/attrite	**Probability to Dropout area 1 incl.**	**Probability to Dropout area 1 excl.**	**Probability of Absence area 1 incl.**	**Probability of Absence area 1 excl.**	**Probability of Attrition area 1 incl.**	**Probability of Attrition area 1 excl.**
T1 x Financial Rewards	**-0.048*** **	**-0.053** **	-0.024	-0.024	-0.089	-0.092
	(0.02)	(0.02)	(0.04)	(0.04)	(0.05)	(0.06)
T1 x Reputation Rewards	**-0.057*** **	**-0.064*** **	-0.032	-0.038	**-0.113*** **	**-0.126*** **
	(0.01)	(0.01)	(0.03)	(0.03)	(0.04)	(0.04)
T2 x Financial Rewards	-0.029	-0.029	-0.030	-0.030	-0.061	-0.060
	(0.03)	(0.03)	(0.02)	(0.03)	(0.05)	(0.05)
T2 x Reputation Rewards	**-0.056*** **	**-0.059*** **	**-0.060** **	**-0.052***	**-0.133*** **	**-0.124*** **
	(0.02)	(0.02)	(0.03)	(0.03)	(0.04)	(0.04)
T1 x No Rewards	**-0.025***	-0.007	0.010	**0.057***	-0.015	0.059
	(0.01)	(0.02)	(0.03)	(0.03)	(0.04)	(0.05)
T2 x No Rewards	**-0.028** **	-0.033	0.022	0.013	-0.013	-0.031
	(0.01)	(0.02)	(0.03)	(0.04)	(0.04)	(0.06)
C x Financial Rewards	**-0.050*** **	**-0.053** **	-0.016	-0.011	-0.086	-0.082
	(0.02)	(0.02)	(0.05)	(0.05)	(0.06)	(0.07)
C x Reputation Rewards	-0.016	-0.012	-0.002	0.007	-0.022	-0.007
	(0.02)	(0.02)	(0.03)	(0.03)	(0.05)	(0.05)
Baseline Mathematics and English score	**-0.021*** **	**-0.025*** **	**-0.046*** **	**-0.036*** **	**-0.072*** **	**-0.066*** **
	(0.00)	(0.01)	(0.01)	(0.01)	(0.01)	(0.01)
X (other variables)	…	…	…	…	…	…
N	7108	5042	7108	5042	7108	5042

Note: T1 stands for within-class comparison, T2 for across-class comparison and C for control group. Robust standard errors adjusted for clustering at school level are in parentheses. Controlled for stratum (performance, grade and area) fixed effects. Compared to pure control group (no feedback and no rewards during entire testing). X stands for set of stratification variables controlled for in the regression – area (four different areas), school performance at national examination and grade level (P6,P7, S1 up to S4). N stands for the number of observations. * significant at 10%; ** significant at 5%; *** significant at 1%

4c. Average treatment effect of within and across class competition interacted with financial and reputational reward treatments on students' performance

The core question of the experiment was whether social comparison (and what type) can motivate students to improve their performance via enhanced competitiveness and whether the effects would differ if the students were additionally rewarded financially or reputationally. Table 9 summarizes the results. First, I present the results of standard OLS. In order to control for non-random attrition, I proceeded with inverse probability-weighted regressions. Inverse probability weighting (IPW) can adjust for confounding factors and selection bias. It assigns a weight to every

student based on their probability to dropout (columns 3 and 4) or probability to attrite (columns 5 and 6) and adjust for that in estimation of the treatment effects. All the scores (baseline as well as endline scores) were normalized with respect to the control group in round 1 in order to express the results in standard deviation.

The results suggest that within class as well as across class competition influence students' performance mildly, causing improvement of 0.064 standard deviations by OLS results and 0.11-0.14 standard deviations after inverted weighting. In general, usage of inverted probability weights pronounces the effect sizes due to controlling for probability to dropout or probability to attrite. Among the real rewards, only financial rewards seem to have positive and significant effect on students' performance. Financial treatment led to 0.17 - 0.23 standard deviation improvement in students' performance. An effect size of 0.2 standard deviations and above is considered as strong effect in education literature. Reputational rewards did not motivate students to improve their performance.

Table 9: THE EFFECTS OF TREATMENTS ON STUDENTS' PERFORMANCE

Students' score from Math and English at the end of the experiment	**OLS area 1 incl.**	**OLS area 1 excl.**	**Weighted Regression area 1 incl.**	**Weighted Regression area 1 excl.**	**Weighted Regression area 1 incl.**	**Weighted Regression area 1 excl.**
Within class comparison	**0.061***	0.059	**0.114****	0.096	**0.097***	**0.098***
(T1)	(0.036)	(0.045)	(0.055)	(0.059)	(0.049)	(0.058)
Across class comparison	0.064	0.061	**0.111***	**0.130****	**0.107****	**0.121****
(T2)	(0.044)	(0.050)	(0.056)	(0.056)	(0.050)	(0.052)
Financial Rewards	**0.187****	**0.189*****	**0.226****	**0.237****	**0.169***	**0.173****
	(0.074)	(0.067)	(0.109)	(0.097)	(0.091)	(0.083)
Reputational Rewards	0.040	0.038	-0.048	-0.024	-0.066	-0.064
	(0.059)	(0.057)	(0.091)	(0.086)	(0.072)	(0.069)
Baseline Math and English	**0.832*****	**0.840*****	**0.759*****	**0.747*****	**0.806*****	**0.795*****
score	(0.015)	(0.017)	(0.017)	(0.014)	(0.021)	(0.021)
X (other variables)	…	…	…	…	…	…
N	5108	3516	5108	3516	5108	3516

Note: Robust standard errors adjusted for clustering at school level are in parentheses. Controlled for stratum (performance, grade and area) fixed effects. Compared to pure control group (no feedback and no rewards during entire testing). X stands for set of stratification variables controlled for in the regression – area (four different areas), school performance at national examination and grade level (P6,P7, S1 up to S4). N stands for the number of observations. * significant at 10%; ** significant at 5%; *** significant at 1%

Once I allow for interaction between treatments, the results confirm the general result, however, it allows for comparison of treatment interactions. Financial rewards positively influenced students' performance no matter what type of social comparison students faced (T1, T2 or C). Pairwise testing of the effects suggests that being part of social comparison group enhances the effect of competing for financial rewards. The effects are again more pronounced after controlling for probabilities to dropout or to attrite. Moreover, direct comparison of the social comparison groups shows significantly stronger effect if students from financially rewarded group faced across class competition and therefore competed as one big class as opposed to within class competition. Reputational rewards have no effects on students' performance.

Table 10: THE EFFECTS OF TREATMENTS ON STUDENTS' PERFORMANCE

Students' score from Math and English at the end of the experiment	**OLS area 1 incl.**	**OLS area 1 excl.**	**Weighted Regression area 1 incl.**	**Weighted Regression area 1 excl.**	**Weighted Regression area 1 incl.**	**Weighted Regression area 1 excl.**
T1_FinancialRewards	**0.257*** **	**0.225** **	**0.317*** **	**0.249** **	**0.261*** **	**0.243** **
	(0.08)	(0.09)	(0.09)	(0.11)	(0.09)	(0.10)
T1_ReputationalRewards	**0.163***	0.150	0.132	0.102	0.098	0.096
	(0.08)	(0.09)	(0.09)	(0.12)	(0.10)	(0.11)
T2_FinancialRewards	**0.267***	0.249	**0.435*** **	**0.375** **	**0.340** **	**0.329** **
	(0.14)	(0.14)	(0.15)	(0.15)	(0.15)	(0.14)
T2_ReputationalRewards	0.141	0.082	0.153	0.070	0.076	0.032
	(0.09)	(0.09)	(0.11)	(0.13)	(0.10)	(0.11)
T1_NoRewards	0.020	-0.046	0.014	-0.088	0.011	-0.084
	(0.06)	(0.11)	(0.08)	(0.15)	(0.08)	(0.13)
T2_NoRewards	0.054	0.040	0.019	0.024	0.063	0.106
	(0.06)	(0.10)	(0.09)	(0.15)	(0.09)	(0.13)
C_FinancialRewards	**0.195** **	**0.175***	0.169	0.152	0.170	0.179
	(0.08)	(0.10)	(0.11)	(0.13)	(0.10)	(0.12)
C_ReputationalRewards	0.066	0.010	0.006	-0.079	-0.039	-0.080
	(0.09)	(0.11)	(0.11)	(0.12)	(0.09)	(0.11)
Baseline Mathematics and	**0.831*** **	**0.838*** **	**0.759*** **	**0.749*** **	**0.807*** **	**0.795*** **
English score	(0.01)	(0.02)	(0.02)	(0.01)	(0.02)	(0.02)
X (other variables)	…	…	…	…	…	…
N	5108	3516	5108	3516	5108	3516

Note: T1 stands for within-class comparison, T2 for across-class comparison and C for control group. Robust standard errors adjusted for clustering at school level are in parentheses. Controlled for stratum (performance, grade and area) fixed effects. Compared to pure control group (no feedback and no rewards during entire testing). X stands for set of stratification variables controlled for in the regression – area (four different areas), school performance at national examination and grade level (P6,P7, S1 up to S4). N stands for the number of observations. * significant at 10%; ** significant at 5%; *** significant at 1%

The effects do not seem to be heterogeneous in terms of students' baseline score; it means that better performing students do not necessarily react to treatment differently compared to poor performing students.

Table 11: THE HETEROGENEITY OF THE EFFECTS OF TREATMENTS ON STUDENTS' PERFORMANCE

Students' score from Math and English at the end of the experiment	**OLS area 1 incl.**	**OLS area 1 excl.**	**Weighted Regression area 1 incl.**	**Weighted Regression area 1 excl.**	**Weighted Regression area 1 incl.**	**Weighted Regression area 1 excl.**
Within class comparison	0.049	0.052	0.055	0.038	0.055	0.054
(T1)	(0.04)	(0.05)	(0.05)	(0.05)	(0.05)	(0.05)
Across class comparison	0.063	0.062	**0.102***	**0.117***	**0.100***	**0.117****
(T2)	(0.05)	(0.05)	(0.06)	(0.06)	(005)	(0.05)
Financial Rewards	**0.208*****	**0.206*****	**0.318*****	**0.311*****	**0.241****	**0.237****
	(0.07)	(0.07)	(0.10)	(0.10)	(0.09)	(0.09)
Reputational Rewards	**0.097***	0.086	**0.127***	**0.137***	0.046	0.040
	(0.05)	(0.06)	(0.07)	(0.07)	(007)	(0.07)
T1 x Baseline Performance	**0.064***	0.064	0.042	0.047	0.064	0.076
	(0.03)	(0.04)	(0.03)	(0.04)	(0.04)	(0.05)
T2_Baseline Performance	0.042	0.043	0.016	-0.011	0.033	0.029
	(0.04)	(0.05)	(0.04)	(0.04)	(0.04)	(0.05)
Financial Rewards x	0.025	0.025	-0.015	-0.025	-0.040	-0.056
Baseline Performance	(0.02)	(0.03)	(0.04)	(0.04)	(0.04)	(0.06)
Reputation Rewards x	0.004	0.006	**-0.085*****	**-0.078****	**-0.070***	-0.078
Baseline Score	(0.04)	(0.04)	(0.03)	(0.04)	(0.04)	(0.06)
Baseline Mathematics and	**0.786*****	**0.790*****	**0.799*****	**0.816*****	**0.806*****	**0.814*****
English score	(0.03)	(0.04)	(0.04)	(0.05)	(0.04)	(0.05)
X (other variables)	…	…	…	…	…	…
N	5108	3516	5108	3516	5108	3516

Note: Robust standard errors adjusted for clustering at school level are in parentheses. Controlled for stratum (performance, grade and area) fixed effects. Compared to pure control group (no feedback and no rewards during entire testing). X stands for set of stratification variables controlled for in the regression – area (four different areas), school performance at national examination and grade level (P6,P7, S1 up to S4). N stands for the number of observations.
* significant at 10%; ** significant at 5%; *** significant at 1%

5. Conclusion

This paper contributes to the current literature by studying the effects of social comparisons on students' performance. It directly allows differentiating between within class social comparison and across class social comparison. The treatment is based on pure feedback provision since students are rewarded symbolically only (students received small report cards carrying the feedback information). No further incentivization was offered at the first stage. The results of existing studies are conflicting, Andrabi, Das and Ijaz-Khwaja (2009) found strongly positive effects, Erickson et al. (2009) found no effect and negative effects in the study of Bandiera (2011). In this case, feedback provision seems to have positive impact on students' performance, ranging from 0.06 to 0.13 standard deviations. The effects become more pronounce once the social comparison is intensified by offering rewards to students. Two types of rewards were offered – financial rewards (2000 Ugandan Shillings) and reputational rewards (winners' names announced in local newspapers). By this design I am able to compare directly the value added of these two types of rewards. While financial rewards seem to motivate all students with or without social comparison treatment to perform better and improve students' performance by 0.25 standard deviations, reputational rewards have no effect on students' performance. Both social comparison treatments have an effect on students' dropout rates, too, ranging from 3.7 per cent to 4.3 per cent. Absence rates do not seem to be influenced by any of the treatments. However, since the absence and dropout rates were not collected during random visits, the results should be interpreted with caution.

References:

Alcazar, L., Rogers, F.H., Chaudhury, N., Hammer, J., Kremer, M. and Muralidharan, K. (2006): Why are teachers absent? Probing service delivery in Peruvian primary schools, International Journal of Educational Research, Vol. 45, Issue 3, p. 117-136

Alicke, M. D. (1985): Global self-evaluation as determined by the desirability and controllability of trait adjectives, Journal of Personality and Social Psychology, 49, p.1621–1630

Andrabi, T., Das J. and Ijaz-Khwaja, A. (2009): Report Cards: The Impact of Providing School and Child Test-scores on Educational Markets, BREAD Working Paper No. 226

Angrist, J. and Lavy, V. (1999): Using Maimonides' rule to estimate the effect of class size on scholastic achievement, The Quarterly Journal of Economics, MIT Press, Vol. 114(2), p. 533-575

Angrist. J. and Lavy, V. (2002): The Effect of High School Matriculation Awards: Evidence from Randomized Trials, NBER Working Papers 9389, National Bureau of Economic Research, Inc.

Bandiera, O., Barankay, I. and Rasul, I. (2004): Relative and Absolute Incentives: Evidence on Worker Productivity, CEPR Discussion Papers 4431, C.E.P.R. Discussion Papers

Bandiera, O., Barankay, I. and Rasul, I. (2009): Social Incentives in the Workplace, IZA Discussion Papers 4190, Institute for the Study of Labor (IZA)

Banerjee, A., Barnerji, R., Duflo, E., Glennerster, R. and Khemani, S. (2008): Pitfalls of Participatory Programs: Evidence from a Randomized Evaluation in Education in India, NBER Working Papers 14311, National Bureau of Economic Research, Inc.

Banerjee, A., Cole, A., Duflo, E. and Linden, L. (2007): Remedying Education: Evidence from Two Randomized Experiments in India, The Quarterly Journal of Economics, MIT Press, Vol. 122(3)

Bettinger, Kremer and Saavedra (2010): Are Educational Vouchers Only Redistributive? The Economic Journal, Vol. 120, Issue 546, p. F204–F228

Bjorkman, M. (2004): Public Funding in the Educational Sector and Its Effect on Test Scores, Working paper, IIES, Stockholm University

Bjorkman, M. and Svensson, J. (2009): Power to the People: Evidence from a Randomized Field Experiment of a Community-Based Monitoring Project in Uganda, CEPR Discussion Papers 6344, C.E.P.R. Discussion Papers

Blimpo (2010): Team Incentives for Education in Developing Countries A Randomized Field Experiment in Benin, Job market paper, New York University

Brown, J. D. (1986): Evaluations of self and others: Self-enhancement biases in social judgments. Social Cognition, 4, p.353–376

Burson, K. A., Larrick, R. P. and Klayman, J. (2006): Skilled or unskilled, but still unaware of it: How perceptions of difficulty drive miscalibration in relative comparisons. Journal of Personality and Social Psychology, 90, p.60–77

Carpenter, J. and Seki, E. (2005): Do Social Preferences Increase Productivity? Field experimental evidence from fishermen in Toyoma Bay, Middlebury College Working Paper Series

Chaudhury, N., Hamer, J., Kremer, M., Muralidharan, K. and Rogers, F.H. (2006): Missing in Action: Teacher and Health Worker Absence in Developing Countries, Journal of Economic Perspectives, Vol. 20, No. 1, p. 91–116

Dijkstra, P., Kuyper, H., van der Werf, G., Buunk, A.P. and van der Zee, Y. (2008): Social comparison in the classroom: a review, Review of educational research, Vol. 78, No. 4, p.828-879

Dunning, D., Johnson, K., Ehrlinger, J. and Kruger, J. (2003): Why people fail to recognize their own competence. Current Directions in Psychological Science, 12, p.83–87

Ehrlinger, J., Johnson, K., Banner, M., Dunning, D. and Kruger, J. (2008): Why the unskilled are unaware: Further explorations of (absent) self-insight among the incompetent, Organizational Behavior and Human Decision Processes 105, p.98–121

Duflo, E. and Hanna, R. (2006): Monitoring Works: Getting Teachers to Come to School, C.E.P.R. Discussion Papers CEPR Discussion Papers: 5426

Falk, A. and Ichino, A. (2006): Clean Evidence on Peer Pressure, Journal of Labor Economics, Vol. 24, Issue 1

Fehr, E. and Fischbacher, U. (2002): Why social preferences matter – the impact of non-selfish motives on competition, cooperation and incentives, Economic Journal 112, p.C1-C33

Glewwe, P., Illias , N. and Kremer, M. (2003): Teacher incentives, NBER Working Paper No. 9671. National Bureau of Economic Research, Inc.

Glewwe, P., Kremer, M. and Moulin, S. (2002): Textbooks and Test Scores: Evidence from a Prospective Evaluation in Kenya. Working paper, Harvard University

Glewwe, P., Kremer, M., Moulin, S. and Zitzewitz, E. (2004): Retrospective vs. Prospective Analyses of School Inputs: The Case of Flip Charts in Kenya, Journal of Development Economics, 74 (1), p.251-268

Hanushek, A., Woessman, L. (2007): The role of education quality for economic growth, World Bank Policy Research Working Paper 4122

Hanushek, A., Woessman, L. (2009): Schooling, cognitive skills, and the Latin American growth puzzle, NBER Working Papers 15066, National Bureau of Economic Research, Inc.

Kelley, C., Heneman, H. and Milanowski, A. (2002): Teacher Motivation and School-Based Performance Awards, Educational Administration Quarterly 38

Krajc, M. (2008): Are the unskilled really that unaware? Understanding seemingly biased self-assessment, CERGE-EI Working Papers, WP373

Krajc, M., Ortmann, A. (2008): Are the unskilled really that unaware? An alternative explanation, Journal of Economic Psychology 29, p.724-738

Kremer, M., Miguel, E. and Thornton, R. (2002): Incentives to Learn, NBER Working Papers 10971, National Bureau of Economic Research, Inc.

Kremer, M., Muralidharan, K., Chaudhury, N., Hammer, J., Halsey, F. (2004): Teacher Absence in India, Working paper, World Bank

Kremer, M. and Chen, D. (2001): An Interim Report on a Teacher Attendance Incentive Program in Kenya, Harvard University, Mimeo

Krueger, I.J. and Funder, D. C. (2004): Toward a balanced social psychology: Causes, consequences, and cures for the problem seeking approach to social behavior and cognition. Behavioral and Brain Sciences, 27, p.313–327

Krueger, I.J., Mueller, A.R. (2002): Unskilled, Unaware, or Both? The Better-Than-Average Heuristic and Statistical Regression Predict Errors in Estimates of Own Performance. Journal of Personality and Social Psychology, 82, p.180-188

Kruger, J. and Dunning, D. (1999): Unskilled and Unaware of It: How difficulties in Recognizing One's Own Incompetence Lead to Inflated Self-Assessments, Journal of Personality and Social Psychology, Vol. 77, No. 6., p.121-1134

Kosfeld, M. and Neckermann, S. (2011): Getting More Work for Nothing? Symbolic Awards and Worker Performance, American Economic Journal: Microeconomics, Vol. 3, Issue 3

Lavy , V. (2002): Paying for Performance: The Effect of Teachers' Financial Incentives on Students' Scholastic Outcomes, CEPR Discussion Papers 3862

Lavy, V. (2009): Performance Pay and Teachers' Effort, Productivity and Grading Ethics, American Economic Review 99, 5

Mas, A. and Moretti, E. (2009): Peers at Work, American Economic Review, Vol. 99, Issue 1

Pandey, P., Goyal, S. and Sundararaman, V. (2009): Community Participation in Public Schools: Impact of Information Campaigns in Three Indian states, Education Economics, vol. 17, no. 3, p. 355–375

Pradhan, M., Suryadarmna, D. and Beatty, A. (2010): Improving Learning Outcomes through Enhancing Community Participation: Evidence from a Randomised Field Experiment in Indonesia,

Amsterdam Institute for International Development, Australian National University, World Bank, Mimeo

Ray, D. (2002): Aspirations, poverty and economic change, Understanding Poverty, 2006, p. 409-443(35)

Reinikka, R. and Svensson, J. (2004): Local Capture: Evidence from a Central Government Transfer Program in Uganda." Quarterly Journal of Economics, 119 (2), p.679-705

Reinikka, R. and Svensson, J.(2005): Fighting Corruption to Improve Schooling: Evidence from a Newspaper Campaign in Uganda, Journal of the European Economic Association, MIT Press, vol. 3(2-3), p. 259-267

Shastry, Gauri Kartini and Leigh Linden (2008): Identifying Agent Discretion: Exaggerating Student Attendance in Response to a Conditional School Nutrition Program, Working paper, Columbia University Department of Economics

Appendix A: BALANCE BETWEEN CONTROL AND TREATMENT GROUPS

Variable	Control	Treatment 1	Treatment 2
School Level:	10	11	10
The number of primary schools	7	7	8
The number of secondary schools			
School Type:			
Public Schools	8	5	6
Private Schools	7	9	8
Community Schools	2	4	4
By Population	2345 (48 groups)	2415 (51 groups)	2371 (51 groups)
By PLE/UCE results	3.175	3.039	3.102
By testing results	21.140	21.363	21.648

Note: min(PLE/UCE)= 1.7397, max(PLE/UCE)= 4.2857, mean(PLE/UCE)=3.1040
Note: min(TR)=8.3125, max(TR)=39.7765, mean(TR)=21.3192, where TR=Testing Results

Appendix B: DIFFERENCES IN BASELINE VARIABLES BETWEEN TREATMENT AND CONTOL GROUPS (Full scale)

	After Math Questionnaire		After English Questionnaire	
Variable	**(T1 – C)**	**(T2 – C)**	**(T1 – C)**	**(T2 – C)**
Question 1: Expected number of points				
0 – 10 points	-0.927	0.036	-0.985	0.315
	(1.79)	(0.97)	(2.52)	(1.29)
11 – 20 points	0.970	1.116	1.375	0.055
	(1.45)	(0.89)	(1.55)	(0.76)
21 – 30 points	0.648	0.301	0.095	0.053
	(1.65)	(0.76)	(1.73)	(0.85)
31 – 40 points	-0.508	0.229	-0.496	0.042
	(1.75)	(1.03)	(1.37)	(0.68)
41 – 50 points	-1.111	0.217	-1.045	-0.018
	(1.86)	(1.13)	(2.08)	(1.12)
51 – 60 points	-2.006	-0.248	0.403	0.260
	(2.33)	(1.26)	(2.21)	(1.18)
61 – 70 points	2.355	0.980	0.325	1.245
	(3.19)	(1.63)	(2.80)	(1.29)
71 – 80 points	3.423	0.170	0.365	1.262
	(4.25)	(1.81)	(2.96)	(1.31)
81 – 90 points	2.419	0.711	1.936	0.556
	(4.86)	(2.25)	(3.94)	(1.85)
91 – 100 points	3.456	-0.333	1.985	-0.537
	(5.64)	(2.80)	(5.38)	(2.22)
Question 2: Subjective effort level				
I did not put any effort	1.496	1.614	-3.452	-0.018
	(2.06)	(1.15)	(2.97)	(1.55)
I put little effort	-0.073	0.334	-0.212	-0.287
	(1.79)	(1.05)	(2.04)	(1.03)
I put some effort	0.164	0.374	-0.500	0.056
	(1.91)	(1.08)	(1.88)	(0.97)
I put a lot of effort	0.059	0.116	0.075	0.461
	(2.65)	(1.27)	(2.27)	(1.22)
I did my absolutely best	-0.464	-0.109	0.616	0.186
	(3.44)	(1.70)	(3.75)	(1.82)

Note: Robust standard errors adjusted for clustering at school level are in parentheses. The table shows the difference in response rate between treatment and control group.
* significant at 10%; ** significant at 5%; *** significant at 1%

DIFFERENCES IN BASELINE VARIABLES BETWEEN TREATMENT AND CONTOL GROUPS (CONT.)
(Full Scale)

	After Math Questionnaire		After English Questionnaire	
Variable	**(T1 – C)**	**(T2 – C)**	**(T1 – C)**	**(T2 – C)**
Question 3: Perceived difficulty				
It was much more difficult	-2.349	-0.861	-3.554	-1.516
	(2.49)	(1.39)	(2.46)	(1.31)
It was more difficult	-1.215	-0.561	0.035	-0.718
	(1.52)	(0.78)	(1.97)	(0.91)
It was of comparable difficulty	0.259	0.689	0.360	1.038
	(2.07)	(1.15)	(2.25)	(1.18)
It was easier	1.464	0.201	-0.147	0.425
	(2.76)	(1.29)	(2.30)	(1.19)
It was much easier	-0.866	0.113	1.006	-0.135
	(3.84)	(1.94)	(3.38)	(1.69)
Question 4: Subjective level of happiness				
Very very happy	0.506	0.403	-0.043	0.251
	(2.89)	(1.42)	(2.79)	(1.32)
Very happy	1.148	0.395	1.139	0.845
	(2.97)	(1.39)	(2.48)	(1.25)
Little happy	-0.889	0.072	-0.183	0.392
	(2.05)	(1.16)	(2.19)	(1.24)
Neutral	0.293	0.172	-0.541	-0.876
	(2.78)	(1.57)	(3.11)	(1.62)
Little unhappy	0.895	1.889	-1.133	-0.111
	(2.02)	(1.27)	(2.63)	(1.44)
Very unhappy	-0.753	-1.223	-0.657	-0.144
	(2.79)	(1.39)	(2.51)	(1.32)
Very very unhappy	-2.363	0.609	-2.749	-0.325
	(2.57)	(1.70)	(2.55)	(1.75)

Note: Robust standard errors adjusted for clustering at school level are in parentheses. Controlled for stratum (performance, grade and area) fixed effects. The table shows the difference in response rate between treatment and control group.
* significant at 10%; ** significant at 5%; *** significant at 1%

Appendix C: OBSERVING THE EFFECTS OF TREATMENTS ON PROBABILITY TO DROPOUT, ABSENCE AND ATTRITION RATES

Dependent variable: probability of dropout/absent/attrite	Probability to Dropout area 1 incl.	Probability to Dropout area 1 excl.	Probability of Absence area 1 incl.	Probability of Absence area 1 excl.	Probability of Attrition area 1 incl.	Probability of Attrition area 1 excl.
Within class social comparison (treatment 1)	**-0.041*** **	**-0.043** **	-0.016	-0.009	**-0.064** **	-0.060
	(0.01)	(0.02)	(0.02)	(0,03)	(0.03)	(0.04)
Across class social comparison (treatment 2)	**-0.037*** **	**-0.041*** **	-0.018	-0.033	**-0.061** **	**-0.081** **
	(0.01)	(0.01)	(0.02)	(0.03)	(0.03)	(0.04)
Financial Rewards	-0.027	-0.030	-0.031	-0.034	-0.066	-0.069
	(0.02)	(0.02)	(0.03)	(0.03)	(0.05)	(0.05)
Reputational Rewards	**-0.033** **	**-0.036** **	-0.033	-0.036	**-0.078***	**-0.082***
	(0.01)	(0.02)	(0.03)	(0.03)	(0.05)	(0.05)
Baseline Mathematics and English score	**-0.021*** **	**-0.025*** **	**-0.047*** **	**-0.038*** **	**-0.073*** **	**-0.069*** **
	(0.00)	(0.01)	(0.01)	(0.01)	(0.01)	(0.01)
Performance	-0.013	**-0.078** **	0.007	**-0.056** **	-0.004	**-0.129*** **
	(0.01)	(0.03)	(0.04)	(0.03)	(0.05)	(0.05)
Area 2	**0.233** **	**0.167***	**0.287*** **	**0.223*** **	**0.490*** **	**0.419*** **
	(0.09)	(0.09)	(0.09)	(0.07)	(0.16)	(0.16)
Area 3	0.105	0.065	-0.004	-0.044	0.105	0.051
	(0.07)	(0.07)	(0.07)	(0.07)	(0.11)	(0.11)
Area 4	0.039		0.043		0.062	
	(0.05)		(0.05)		(0.07)	
Area 2 x Performance	-0.017	0.042	0.026	0.074	0.001	0.106
	(0.03)	(0.05)	(0.06)	(0.06)	(0.11)	(0.12)
Area 3 x Performance	-0.031	0.025	-0.065	-0.017	-0.109	-0.006
	(0.02)	(0.05)	(0.05)	(0.05)	(0.07)	(0.08)
Area 4 x Performance	**-0.046** **		-0.050		**-0.103** **	
	(0.02)		(0.04)		(0.05)	
P6	**-0.054*** **	**-0.062** **	**-0.117*** **	**-0.117*** **	**-0.187*** **	**0.200*** **
	(0.02)	(0.02)	(0.02)	(0.03)	(0.03)	(0.03)
P7	**-0.079*** **	**-0.094*** **	**-0.150*** **	**-0.164*** **	**-0.245*** **	**-0.287*** **
	(0.01)	(0.02)	(0.02)	(0.02)	(0.02)	(0.02)
S1	-0.020	-0.021	0.017	0.023	-0.002	0.005
	(0.02)	(0.03)	(0.05)	(0.07)	(0.04)	(0.06)
S2	-0.017	-0.010	-0.049	-0.055	**-0.073** **	-0.072
	(0.01)	(0.02)	(0.03)	(0.04)	(0.03)	(0.04)
S3	-0.007	-0.018	-0.045	**-0.049***	**-0.057***	**-0.075***
	(0.03)	(0.04)	(0.02)	(0.03)	(0.03)	(0.04)
N	7108	5042	7108	5042	7108	5042

Note: Robust standard errors adjusted for clustering at school level are in parentheses. Controlled for stratum (performance, grade and area) fixed effects. X stands for set of stratification variables controlled for in the regression – area (four different areas), school performance at national examination and grade level (P6 and P7 of primary schools, S1 up to S4 of secondary schools) and the interaction terms of area and school performance. N stands for the number of observations. * significant at 10%; ** significant at 5%; *** significant at 1%