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

DAGMARA CELIK KATRENIAK§

**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, quality downturn. A large number of interventions have been implemented with the aim of lowering absenteeism and improving students' performance. One possible channel is the provision of feedback about the subject's position in the group. Subjects also seem to improve if evaluated in groups and/or if provided with incentives, such as financial and reputational rewards. This paper contributes to the discussion in five aspects by implementing two types of social comparative feedback regimes - within and across-class group comparisons and two incentive regimes. First, it studies the effects of comparative feedback on students' group performance without further incentivization. Second, it helps to understand the value added of financial and reputational rewards introduced into a social comparison framework. Third, it provides us with further evidence on gender differences in responding to incentives. Fourth, it contemplates the effects of incentives on additional outcomes, such as happiness and stress. Finally, the paper contributes to the scarce literature on the provision of incentives in developing country.. Disaggregation of the treatment interaction reveals that both girls and boys react similarly to the interactions of social comparison treatment with rewards, the channels are, however, different. While girls' improvement is driven mainly by the comparative feedback, boys react to rewards only with no added value of feedback provision. The results are heterogeneous also with respect to students' initial ability distribution – mostly students above median respond positively.

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

---

* This research was supported by GA UK grant No. 338911, and by GDN grant No.60. All errors are mine.
¶ CERGE-EI (a joint workplace of the Center for Economic Research and Graduate Education of Charles University, and the Economic Institute of the Academy of Sciences of the Czech Republic, v.v.i.), Politickych veznu 7, 11121, Prague 1, Czech Republic, and National Research University Higher School of Economics, Myasnitskaya 20, 101000, Moscow, Russia. Email: dkatreni@cerge-ei.cz.

## 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 many approaches to improving education quality, one venue that has been explored is 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 (Festinger, 1954; the founder of the social comparison theory). On the contrary, comparison enables a child to find his/her relative position within a particular group which can lead via enhanced competitiveness to an increase in effort and performance improvement. **Feedback provision**, as a way to inform subjects about their absolute or relative standing, has been analyzed in different environments and has delivered conflicting results. 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. Azmat and Iriberri (2010) informed high school students about their relative standing and helped this way to improve students' grades by 5 per cent. Erickson et al.. (2009) did not find any effect of feedback of any intensity to influence employee's performance. On the contrary, Bandiera et al. (2011) finds negative effects of feedback provision. Workers in this 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 effect of feedback, however, depends on who the subjects are compared with, how they are compared and whether they are rewarded for their performance. Are subjects compared individually or in groups? Are groups constructed exogenously or endogenously? Are subjects

compared to others within their ability limits or to much better performers? Are subjects rewarded for their improvements?

Students face social comparison in their classrooms on a daily basis and it can strongly influence their self-esteem and their performance (Dijskstra et al., 2008). It is therefore important to understand with whom to optimally compare the students. If students are **compared to** the ones slightly better, their effort and performance tend to increase. Performance and effort decrease if the comparison target is too far from a student's ability (Ray, 2003). Students can be compared **individually or in groups**. Group's outcome depends on each member's contribution and may foster mutual help (Slavin, 1984) as well as positive peer effects (Hoxby, 2000; Sacerdote, 2001; Duflo et al., 2011). Groups can be formed **endogenously** (e.g., by students themselves based on friendship) **or exogenously** (Blimpo, 2014) and they can be exposed to competition. In some studies, the effects of interventions are more pronounced if students are involved in **tournaments** (Eriksson et al., 2009 ; Bigoni et al., 2010). Blimpo (2014) tested three incentive schemes – financial rewards for reaching a performance target by individuals (treatment 1), by a group (treatment 2) or by a group involved in a tournament (treatment 3). Groups were formed randomly. All treatments (with or without competition) resulted in positive improvement in students' performance, increased by 0.27 to 0.34 standard deviations.

As in Blimpo (2014), students have been often incentivized to improve their performance. The most commonly used incentives are financial rewards (Fryer, 2010; Angrist et al. 2002 and 2006; Kremer et al., 2002; Bettinger et al., 2010, etc.). However, it is often the case that only students from the second quartile of initial ability distribution react positively to financial rewards leaving other quartile groups unchanged (Bandiera et al., 2012). Financial rewards are often not cost-effective. Therefore further interest lies in studying the effects of **non-pecuniary incentives,** especially status or social recognition rewards. For example, Kosfeld and 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.

Drawback of reward provisions is that they may crowd out **intrinsic motivations** of subjects and decrease their future performances (Deci, 1971). Responses to rewards seem to be gender-sensitive. While girls seem to significantly improve if they are offered financial rewards (Angrist et al., 2009), their responsiveness is lower once involved in competition (Gneezy et al., 2003, Niederle and Vesterlund, 2007).

In the first part of my intervention, I study whether the provision of comparative feedback about group outcomes, a pure information incentive without any rewards, can increase students' effort and lead to performance improvement. Groups are of two types – small groups of 3 to 4 students within class (treatment 1, within class competition) and bigger groups containing all students in the class (treatment 2, across class competition). Students are tested repeatedly during an academic year and receive pure information feedback three (treatment 1) to four times (treatment 2). In order to see the value added of rewards additionally introduced and their interactions with social comparison, I orthogonally re-randomize the sample before the final school visit and offer financial rewards to 15 per cent of the best performing and 15 per cent of the most improving groups. The additional reward treatments together with the initial social comparison treatments divide my sample into nine groups – within class competition with financial, reputational or no rewards, across class competition with financial, reputational or no rewards and control groups without being exposed to an social comparison but reward financially, reputationally and finally pure control group (no feedback, no rewards). Groups are formed exogenously. However, by means of additional questionnaires on students' four best friends and a class idol I can control for the level of friendship within groups. Additional outcomes such as happiness and stress and their responses to the treatments are also analyzed.

The predictions of the effects of my interventions based on existing literature are controversial. Evaluation of students in groups should push via enhanced cooperation within groups to group average improvements. If the group is, however, big enough, free riding behavior may prevail and result in heterogeneity within the group outcomes. Informing students about the position of their group could lead to improvements in performances via enhanced competition or demotivate students with negative attitude toward competition. The effect potentially depends on group composition (gender, friendship or ability composition) and how much worse the group in the group ability distribution. Students included in both financial as well as reputational reward treatments are expected to improve their scores, at least the ones in the second quartile of ability distribution.

The treatments effect students' participation as well as their performance in the final examination. The effect is driven mainly by improvements in girl's performance in English, leaving Mathematic score without a significant change. Ordinary least squares estimates of the overall effects of the social comparison feedback suggest 0.2 to 0.25 standard deviation improvement in English scores and 0.3 to 0.44 standard deviation improvement if rewarded reputationally or financially. Once the effects are decomposed by interactions of all the treatments one can see the opposing effects.

The paper is organized as follows. Section 2 summarizes the methodology; it describes randomized control trials and randomization scheme, and the experimental design. Section 3 summarizes the logistics and the timeline of the experiment, and overall sample. Section 4 summarizes the results of the intervention. Section 5 concludes the main findings.

## 2. Experimental design and randomization

To evaluate the effect of the intervention, I designed a Randomized Control Trial (RCT) experiment. If the randomization is done properly with a high enough number of random draws, it ensures a balance between control and treatment groups in expectations in terms of observables as

well as unobservables. Therefore there should be neither selection bias nor confounding factors to spoil the program evaluation (Duflo et al., 2006). In order to increase the balance between control and treatment groups, the sample is stratified along three dimensions – school's location (the sample was divided into four areas differing in the level of remoteness), average school performance in national examination (above average or below average) and students' level (grade 6 and 7 of primary education and grades 1 up to 4 of secondary education) [1,2]. Stratification divided my sample into 48 stratas. Within each strata, I randomized the sample into treatment and control groups.

The randomization was done in two stages (as shown in Figure 1). First, after the 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, I divided classes of the treatment schools 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

[1] Every year students of P7 in primary schools and S4 in secondary schools take the national 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.

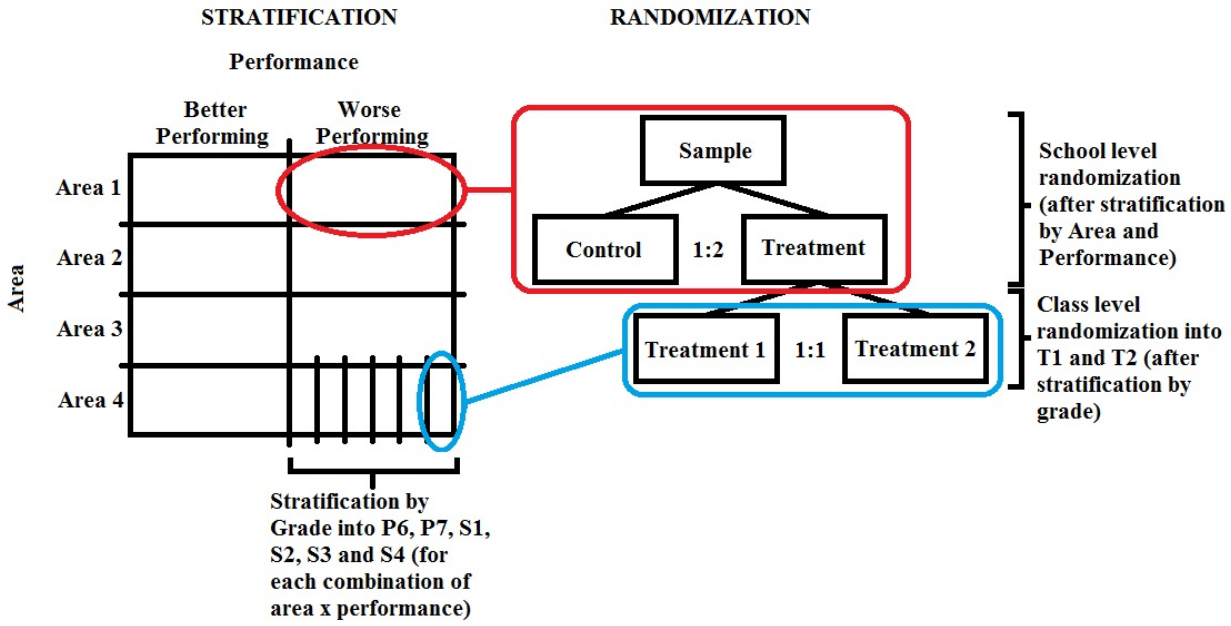
[2] Uganda introduced Universal Primary Education (UPE) for all in 1997, allowing up to four students to go to school for free. Later it was extended to all children. Primary education is a seven-year program and for successful completion students need to pass the national Primary Leaving Exam (PLE) 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.

same treatment. However, it was not feasible in my case due to budget, the sample size as well as 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.

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. In other words, group averages were taken into account when comparing the students' performance. Students in treatment 2 were evaluated as a whole class (using class average) and were compared to other classes of the same grade in different schools.

**Figure 1: Stratification and randomization scheme**




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, how his group-mates scored 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-mates) 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., they were not given information about their classmates) and the position of their class compared to other classes in the treatment 2. The positions in both treatments were emphasized on a rank-order graph, too (see also Appendix). 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 of such cross-cutting design was to observe whether introduction of additional rewards could enhance students' performance, especially if interacted with the treatments T1 and T2 (see also Figure 2). 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.

**Figure 2: Orthogonal randomization of the sample into reward treatments**

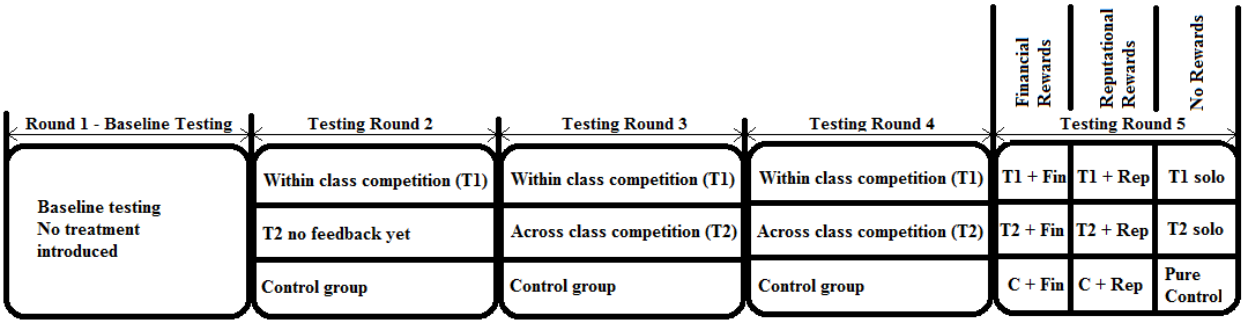



**Table 1: Qualification criteria for winning the rewards**

| | **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.

## 3. Timing, Logistics 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 during 2012 academic year 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 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 ready to help students. Each team member was obliged to answer and clarify any questions students raised in the language most convenient to them (English or Luganda). 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 (suburb of 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. According to UCDT representatives, students were 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.

Students were tested twice per term, which means approximately every one and half month. 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 (two males and two females) and each of them visited one class per school. The agenda of each visit was similar. After we entered the class, treated students received their treatment (social comparison feedback), control students started immediately with questionnaires. 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 would 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.

After students understood the content of the report cards, they were given "Before Math questionnaire"[3], which was followed with Math examination that lasted 30 minutes. Once the time was over, students answered "After Math Before English questionnaire"[4], filled English exam in the subsequent 20 minutes and finally they filled "After English questionnaire"[5]. In order to ensure transparency, I used own constructed tests. Teachers were allowed to be present in the classrooms. 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. During the duration of the experiment there was one attempt of a teacher to help a student (control group school), however the responsible team member informed me on time and teacher was asked not to do so again. Apart from After Math and After English questionnaire, I collected information also about students' aspirations, happiness based on Subjective Happiness

[3] [5] [6] 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.
No before-Math and before-English questionnaires were collected during the baseline survey since students saw the examinations for the first time.

Scale (Lyubomirsky and Lepper, 1997) and their stress level based on Perceived Stress Scale (Cohen, Kamarck and Mermelstein, 1983). This extensive questionnaire was collected at the beginning and at the end of the study. Responses to the questions help me to extend the number of tests for (im)balances between treatment and control group and allow me to observe the effect of the treatment on other outcomes (aspirations, stress and happiness) rather than commonly used Math and English scores.

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; Qualification criteria differed based on initial randomization (T1,T2,C).

## 4. Results

### 4a. Randomization Balance

Table 3 provides a summary of students' mean characteristics and their (im)balances across social comparison treatment and control groups (T1, T2 and C). It disaggregates data on students' performance scores from Mathematics and English (Part A), on students' demographic features, attrition and characteristics from questionnaires (Parts B and C). The balance comparison between reward and no-reward groups based on students' baseline performances can be found in Table 4.

Students achieved on average 11 points from Mathematics (out of 50) and between 11.5 to 12 points from English (out of 50). The performance is not statistically different across social comparison treatment and control groups, suggesting that the randomization divided students into similar groups in terms of their average performance. Students are similar in terms of their characteristics, too. In all three groups students overestimated their performance. While they expected their score from Mathematics to be between 16 to 20 points, in English they felt more confident and expected the total score of 26 to 30 points. Most of the students found both examinations easier compared to their regular examination at school and exerted "a lot of effort" into answering the exam questions (based on 5-likert scale, 1 representing no effort, 5 the maximum effort). According to the 7-likert subjective happiness scale, students were on average "little happy". However students' most frequent response in both cases was that they felt "very happy"(due to the skewed distribution of subjective happiness to the right).

Table 3 also provides information about students' average aspirations, perceived happiness and subjective stress. On average 76 per cent of students prefer educational activity over having rest if they had an extra hour every day, 70 per cent of students would choose educational activity over working for money and approximately 55 per cent of students would work instead of having rest. The average age of the students is 17. In all groups there are more girls than boys participating and

the average class size is 55 students. The last two statistics are based on the restricted sample. Restricted sample differs from overall sample by one school, which went through difficult times during the academic year 2012, exchanged its headmasters twice and it influenced students high turnover. This school caused close to 4 percent of overall attrition of control group schools. Therefore in Table 3, I present statistics for both overall and restricted sample.

Comparison of data on students' performance, demographics and students' responses to questions suggests that randomization divided the sample into groups that are in expectations similar. Significant differences can be observed between treatment 2 and control group, indicating that students in treatment 2 were slightly more stressed, slightly less happy and exerted slightly more effort compared to the control group. If the covariates are correlated with students' performance, such imbalance could bias the estimation of the treatment effect of the intervention (Firpo et al., 2014). One can expect some imbalances between treatment and control groups may occur purely by chance - as the number of balance tests goes up, the probability to reject zero hypothesis of no difference between treatment and control group goes up too. In my case, treatment and control groups differ significantly in less than 5% of all cases (besides attrition, which I discuss in the next section).

**4b. Attrition**

High drop-out and absence rates are common features of students in developing countries and it is not an exception in my data. There are several reasons. Some students did not have money to pay the school fees and decided to change schools to avoid repaying their debt, others changed their school because of family reasons (family moved to different area, they were sent to live with other family members, etc.), some completely dropped out of school, some just registered as new students and some of the students passed away. Due to the constraints of the experiment, all participation data are based on our visits only (it means that no random visits were organized).

Table 3: COMPARISON OF MEAN CHARACTERISTICS OF STUDENTS' IN TREATMENT AND CONTROL GROUPS

| | Means | | | Mean Differences | | Joint P-value |
|---|---|---|---|---|---|---|
| | T1 | T2 | Control | (T1 – C) | (T2 – C) | |
| A. STUDENTS PERFORMANCE – ROUND 1 – BASELINE SURVEY | | | | | | |
| Mathematics | 11.015 | 11.198 | 11.092 | -0.077 | 0.106 | 0.183 |
| | | | | (0.99) | (0.96) | |
| English | 11.551 | 11.927 | 11.477 | 0.074 | 0.450 | 0.699 |
| | | | | (1.53) | (1.72) | |
| Sum Mathematics + English | 22.566 | 23.125 | 22.569 | -0.003 | 0.556 | 0.423 |
| | | | | (2.30) | (2.43) | |
| B. QUESTIONNAIRES | | | | | | |
| **B.1 After Math questionnaire** | | | | | | |
| Q1: Expected number of points | 4.331 | 4.537 | 4.551 | -0.221 | -0.151 | 0.299 |
| [min 1, max 10] | | | | (0.150) | (0.145) | |
| Q2: Subjective effort level | 3.447 | 3.525 | 3.504 | -0.057 | 0.021 | 0.298 |
| [min 1, max 5] | | | | (0.053) | (0.052) | |
| Q3: Perceived difficulty | 3.341 | 3.494 | 3.423 | -0.082 | 0.072 | 0.030 |
| [min 1, max 5] | | | | (0.053) | (0.052) | |
| Q4: Subjective level of happiness | 3.319 | 3.253 | 3.184 | 0.135 | 0.069 | 0.343 |
| [min 1, max 7] | | | | (0.092) | (0.094) | |
| **B.2 After English questionnaire** | | | | | | |
| Q1: Expected number of points | 5.715 | 5.757 | 5.796 | -0.081 | -0.039 | 0.879 |
| [min 1, max 10] | | | | (0.161) | (0.144) | |
| Q2: Subjective effort level | 3.547 | 3.627 | 3.553 | -0.006 | **0.074*** | 0.141 |
| [min 1, max 5] | | | | (0.046) | **(0.044)** | |
| Q3: Perceived difficulty | 3.644 | 3.644 | 3.677 | -0.033 | -0.033 | 0.752 |
| [min 1, max 5] | | | | (0.052) | (0.049) | |
| Q4: Subjective level of happiness | 2.950 | 2.904 | 2.856 | 0.094 | 0.048 | 0.534 |
| [min 1, max 7] | | | | (0.084) | (0.086) | |
| **B.3 Aspiration questionnaire** | | | | | | |
| Aspirations | | | | | | |
| Education over Relax | 3.833 | 3.756 | 3.778 | 0.056 | -0.021 | 0.269 |
| [min 1, max 5] | | | | (0.049) | (0.049) | |
| Education over Work | 3.538 | 3.496 | 3.477 | 0.060 | 0.019 | 0.526 |
| [min 1, max 5] | | | | (0.057) | (0.059) | |
| Work over Relax | 2.766 | 2.701 | 2.803 | -0.037 | -0.102 | 0.524 |
| [min 1, max 5] | | | | (0.094) | (0.090) | |
| Perceived happiness scale | 11.479 | 11.653 | 11.223 | 0.256 | **0.429**** | 0.155 |
| [min 4, max 28] | | | | (0.231) | **(0.222)** | |
| Perceived stress | 6.018 | 6.352 | 5.756 | 0.262 | **0.595***** | 0.000 |
| [min 0, max 16] | | | | (0.164) | **(0.142)** | |

The main concern in most project evaluations is whether the attrition of subjects is random or whether there is a systematic difference between the attrition from the treatment group compared to the control group caused by the intervention itself. The measure of attrition is based solely on students' participation in the first and the last testing round (I will call it attrition). The attrition rate of students in treated classes is by 10-11 per cent lower compared to the control group (6-7 per cent in the restricted sample) and it is statistically significant.

Table 3: COMPARISON OF MEAN CHARACTERISTICS OF STUDENTS' IN TREATMENT AND CONTROL GROUPS (Continued)

| | Means | | | Mean Differences | | Joint P-value |
|---|---|---|---|---|---|---|
| | T1 | T2 | Control | (T1 – C) | (T2 – C) | |
| C. OTHER (continued) | | | | | | |
| **C.1 Attrition rates** | | | | | | |
| All schools | 0.359 | 0.346 | 0.454 | **-0.095*** ** | **-0.108*** ** | 0.002 |
| | | | | **(0.034)** | **(0.033)** | |
| Restricted sample# | 0.358 | 0.348 | 0.417 | **-0.059*** | **-0.069** ** | 0.041 |
| | | | | **(0.030)** | **(0.029)** | |
| **C.2 Alwayscomers** | | | | | | |
| All schools | 0.202 | 0.186 | 0.082 | **0.121*** ** | **0.104*** ** | 0.000 |
| | | | | **(0.033)** | **(0.104)** | |
| Restricted sample# | 0.207 | 0.188 | 0.110 | **0.097*** ** | **0.077** ** | 0.008 |
| | | | | **(0.033)** | **(0.031)** | |
| **C.3 Age** | 17.058 | 17.048 | 16.999 | 0.059 | 0.049 | 0.737 |
| | | | | (0.079) | (0.078) | |
| **C.6 Gender** | | | | | | |
| All schools | 0.534 | 0.512 | 0.508 | **0.025*** | 0.004 | 0.192 |
| | | | | **(0.015)** | (0.015) | |
| Restricted sample# | 0.548 | 0.524 | 0.533 | 0.015 | -0.009 | 0.277 |
| | | | | (0.015) | (0.015) | |
| **C.4 Class size** | | | | | | |
| All schools | 52.26 | 56.42 | 60.00 | **-7.741*** | -3.581 | 0.146 |
| | | | | **(4.045)** | (4.672) | |
| Restricted sample# | 52.15 | 56.56 | 55.14 | -2.985 | 1.428 | 0.489 |
| | | | | (3.988) | (4.651) | |

Attrition rate is defined as the rate of students missing in the last testing round conditional on student's participation in the baseline testing. 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%

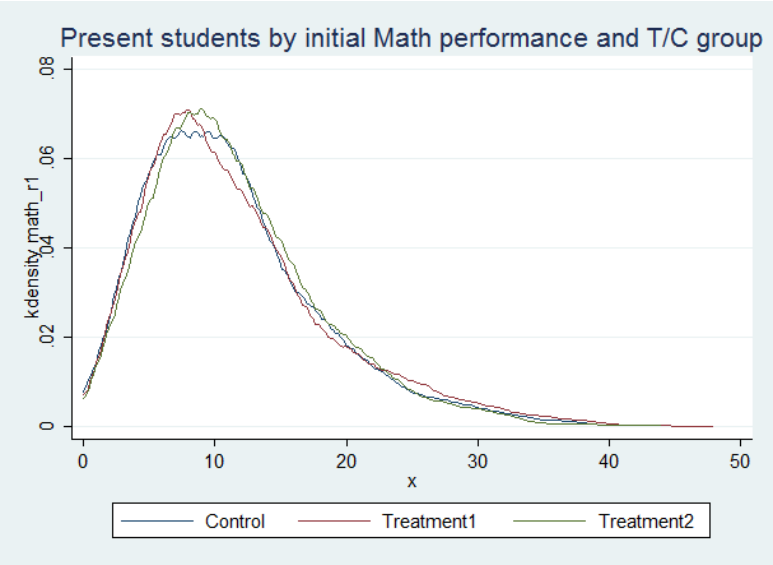
Table 4: DIFFERENCES IN MEANS BETWEEN REWARD TREATMENT AND CONTROL GROUPS

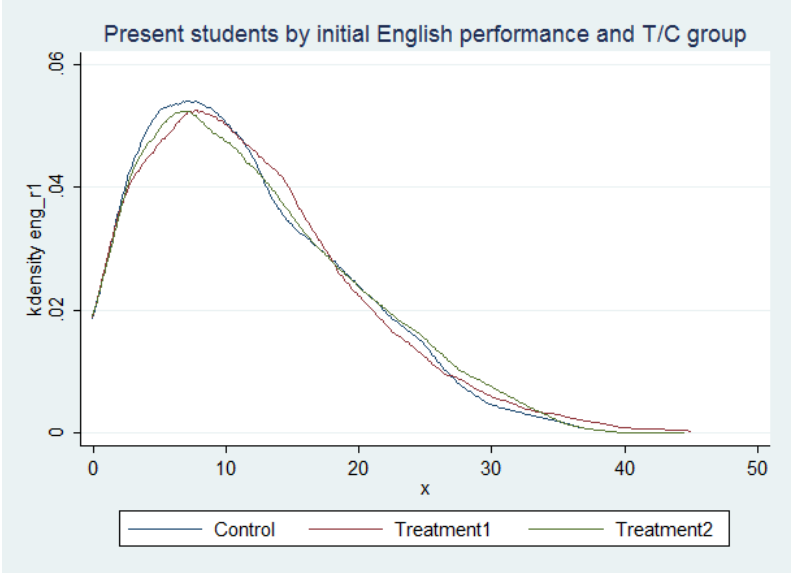
| Variable | FinRew mean | RepRew mean | No Rewards | (Fin – No) std.err. | (Rep – No) std.err. |
|---|---|---|---|---|---|
| Mathematics | 10.038 | 11.200 | 10.231 | -0.193 | 0.969 |
| | | | | (0.94) | (0.88) |
| English | 11.039 | 11.215 | 10.151 | 0.889 | 1.064 |
| | | | | (1.75) | (2.11) |
| Sum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

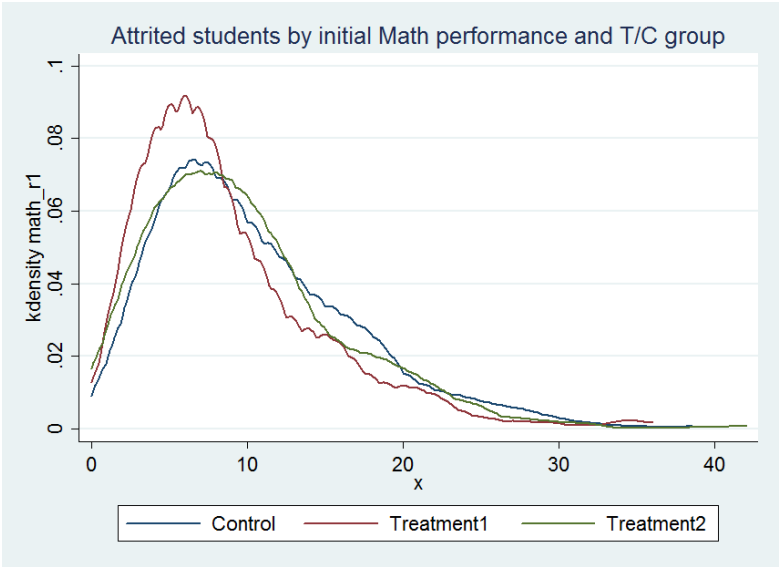
FinRew stands for financially rewarded group, RepRew for reputationally rewarded group and No Rewards represents control group with no rewards. Robust standard errors adjusted for clustering at school level are in parentheses. * significant at 10%; ** significant at 5%; *** significant at 1%

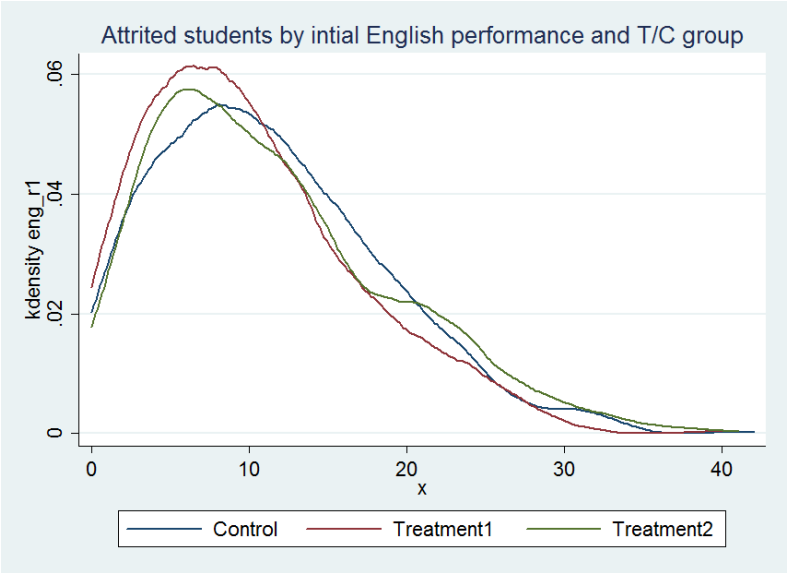
### 4c. Who the attrited students are? Random versus non-random attrition

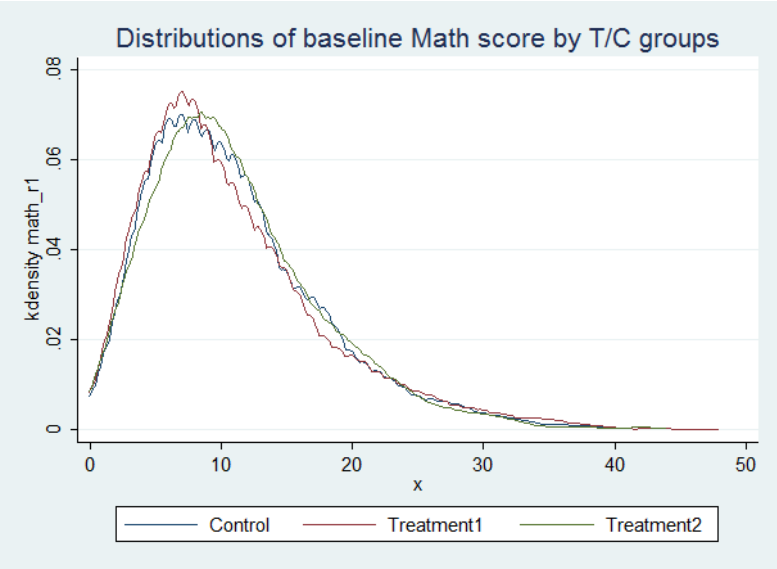
The treatment influenced probability to be always present during our visits and probability to attrite. So in absolute numbers there are less students who dropout from treated classes compared to the control classes and more cases when students from the treatment group attended all five testing rounds compared to students from the control group. Besides the differences in the number of attrited students, students who dropped from T1 were worse in terms of their initial performance compared to students from T2 or the control group. That might re-introduce a bias if treated students who are present during the final testing round are systematically different compared to control-group students. As shown in Table 6, this is not the case of this project. The distribution of students who stayed in either of the treatment groups (based on their initial performance) is not statistically different from the distribution of initial abilities of students from the treatment group. In other words, before as well as after the treatment the composition of students in terms of their initial ability is on average the same. In such case OLS estimate should provide unbiased estimates of the treatment effects. Nevertheless, I will use inverse probability weights and imputation methods to check the stability of the results (for further details see section 4f).


Present students by initial Math performance and T/C group
kdensity math_r1
0
.02
.04
.06
.08
0
10
20
30
40
50
x
Control
Treatment1
Treatment2


Present students by initial English performance and T/C group
kdensity eng_r1
0
.02
.04
.06
0
10
20
30
40
50
x
Control
Treatment1
Treatment2


Attrited students by initial Math performance and T/C group
kdensity math_r1
0
.02
.04
.06
.08
.1
0
10
20
30
40
x
Control
Treatment1
Treatment2


Attrited students by intial English performance and T/C group
kdensity eng_r1
0
.02
.04
.06
0
10
20
30
40
x
Control
Treatment1
Treatment2



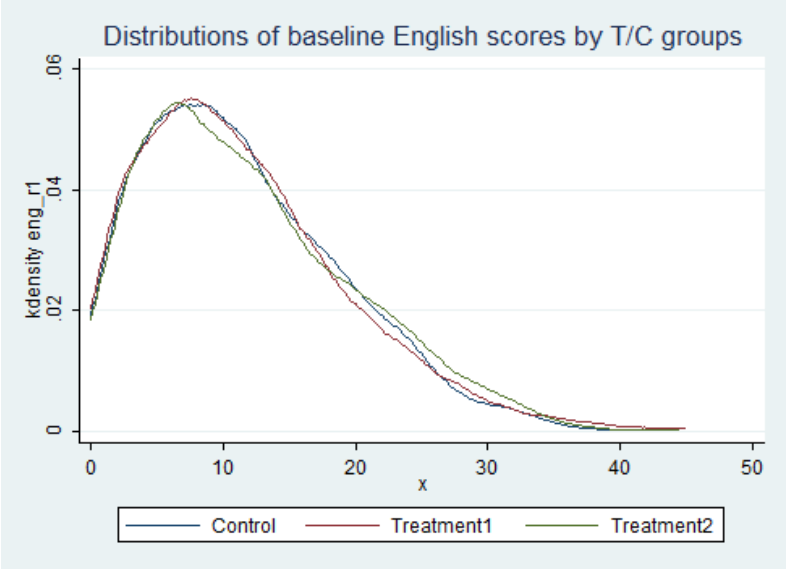




Table 5: Testing of differences in distributions of students who attrited and students who stayed, by T/C group
Ksmirnov test on equality of distributions, p-values presented

| | **Baseline differfences** | | **Students who attrited** | | **Students who stayed** | | **Alwayspresent students** | |
|---|---|---|---|---|---|---|---|---|
| | **(T1 – C)** | **(T2 – C)** | **(T1 – C)** | **(T2 – C)** | **(T1 – C)** | **(T2 – C)** | **(T1 – C)** | **(T2 – C)** |
| A. STUDENTS PERFORMANCE – ROUND 1 – BASELINE SURVEY | | | | | | | | |
| Mathematics | 0.123 | 0.274 | 0.000 | 0.158 | 0.752 | 0.192 | 0.677 | 0.958 |
| English | 0.952 | 0.168 | 0.003 | 0.546 | 0.230 | 0.282 | 0.211 | 0.840 |

Table 6: Comparisons of students based on their attrition/attendance status in terms of their baseline characteristics

| | Baseline differences | | Students who attrited | | Students who stayed | | Alwayspresent students | |
|---|---|---|---|---|---|---|---|---|
| | (T1 – C) | (T2 – C) | (T1 – C) | (T2 – C) | (T1 – C) | (T2 – C) | (T1 – C) | (T2 – C) |
| **B.1 After Math questionnaire** | | | | | | | | |
| Q1: Expected number of points | -0.221 | -0.151 | **-0.332**** | -0.090 | -0.186 | -0.001 | 0.088 | 0.109 |
| [min 1, max 10] | (0.150) | (0.145) | **(0.155)** | (0.171) | (0.171) | (0.160) | (0.194) | (0.184) |
| Q2: Subjective effort level | -0.057 | 0.021 | -0.091 | 0.044 | -0.041 | 0.011 | -0.019 | 0.003 |
| [min 1, max 5] | (0.053) | (0.052) | (0.061) | (0.058) | (0.061) | (0.062) | (0.067) | (0.067) |
| Q3: Perceived difficulty | -0.082 | 0.072 | **-0.128** | **0.100§** | -0.083 | 0.0479 | -0.032 | 0.071 |
| [min 1, max 5] | (0.053) | (0.052) | **(0.077)*** | **(0.069)** | (0.058) | (0.060) | (0.061) | (0.068) |
| Q4: Subjective level of happiness | 0.135 | 0.069 | 0.135 | 0.033 | **0.182*** | 0.132 | **0.163§** | 0.123 |
| [min 1, max 7] | (0.092) | (0.094) | (0.116) | (0.113) | **(0.103)** | (0.101) | **(0.104)** | (0.108) |
| **B.2 After English questionnaire** | | | | | | | | |
| Q1: Expected number of points | -0.081 | -0.039 | -0.294§ | -0.253§ | -0.002 | 0.015 | 0.170 | 0.078 |
| [min 1, max 10] | (0.161) | (0.144) | (0.184) | (0.175) | (0.181) | (0.156) | (0.194) | (0.168) |
| Q2: Subjective effort level | -0.006 | **0.074*** | -0.079 | 0.066 | 0.021 | 0.069 | 0.041 | 0.058 |
| [min 1, max 5] | (0.046) | **(0.044)** | (0.062) | (0.058) | (0.049) | (0.050) | (0.060) | (0.063) |
| Q3: Perceived difficulty | -0.033 | -0.033 | -0.079 | -0.101 | -0.026 | -0.028 | -0.015 | -0.027 |
| [min 1, max 5] | (0.052) | (0.049) | (0.068) | (0.074) | (0.058) | (0.055) | (0.063) | (0.063) |
| Q4: Subjective level of happiness | 0.094 | 0.048 | **0.188*** | 0.153 | 0.105 | 0.065 | 0.049 | 0.072 |
| [min 1, max 7] | (0.084) | (0.086) | **(0.104)** | (0.114) | (0.092) | (0.091) | (0.092) | (0.094) |
| **B.3 Aspiration questionnaire** | | | | | | | | |
| Education over Relax | 0.056 | -0.021 | 0.011 | 0.084 | 0.063 | -0.024 | 0.077 | 0.070 |
| [min 1, max 5] | (0.049) | (0.049) | (0.077) | (0.072) | (0.057) | (0.055) | (0.056) | (0.076) |
| Education over Work | 0.060 | 0.019 | 0.103 | 0.056 | 0.077 | 0.056 | 0.070 | 0.042 |
| [min 1, max 5] | (0.057) | (0.059) | (0.073) | (0.095) | (0.068) | (0.068) | (0.076) | (0.073) |
| Work over Relax | -0.037 | -0.102 | -0.002 | -0.158 | -0.040 | -0.089 | -0.021 | -0.078 |
| [min 1, max 5] | (0.094) | (0.090) | (0.089) | (0.112) | (0.113) | (0.099) | (0.116) | (0.104) |
| Perceived happiness scale | 0.256 | **0.429**** | **0.499§** | 0.459 | 0.209 | **0.399*** | 0.230 | **0.469*** |
| [min 4, max 28] | (0.231) | **(0.222)** | **(0.325)** | (0.349) | (0.244) | **(0.223)** | (0.257) | **(0.241)** |
| Perceived stress | 0.262 | **0.595***** | **0.451**** | **0.381**** | **0.298§** | **0.734***** | **0.355*** | **0.742***** |
| [min 0, max 16] | (0.164) | **(0.142)** | **(0.199)** | **(0.191)** | **(0.199)** | **(0.168)** | **(0.209)** | **(0.172)** |

### 4d. Average treatment effect and its heterogeneity

The core question of the experiment is 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 with financial or reputational rewards. Tables 7a, 7b and 7c summarize the results for Math and English separately. All the scores (baseline as well as endline scores) presented in tables were normalized with respect to the control group in round 1 in respective stratas in order to express the results in standard deviations. While in tables 7a and 7b I show linear combinations of the effects (i.e. overall effects of the treatments), in table 7c I disentangle different channels corresponding to treatment interactions.

The treatment effects are subject-specific. While students exposed to feedback improve only in Mathematics by approximately 0.1 standard deviation (leaving English scores unchanged), rewards have persistent impact on both subjects. Possible explanation is that un-incentivized students who only receive feedback build their expectations on the performance in the exam they are going to undertake (which is Mathematics in my case[6]) and then once they confront the reality (after fulfilling the task) they feel disappointment and they give up. In that case feedback provision would be very sensitive to the gap between true and expected abilities and students' over- or under-confidence. This is not the case in case of financial and reputation rewards, which lead to 0.11 to 0.16 standard deviation improvement in both Math and English. Such pattern can be seen also in the decomposition of the treatment by treatment interactions (Table 7c) as well as by gender decomposition (Tables 8a, b and c).

Girls and boys react to treatments differently. When taken overall treatment effects into account, boys seem to respond to rewards only, while girls improve in response to the feedback provision (see Tables 8a and 8b for overall treatment effects and 8c for the treatment interactions).

[6] The order of exams was not randomized, on contrary, Mathematics was always followed by English exam.

Table 7a: OLS ESTIMATES OF THE EFFECTS OF DIFFERENT MOTIVATION SCHEMES ON STUDENTS' PERFORMANCE IN MATHEMATICS

| **Dependent variable: Math and English score** | **MATHEMATICS** | | | | | |
|---|---|---|---|---|---|---|
| | **Pure FB** | **Pure FB** | **Pure Rewards** | **Pure Rewards** | **Mix FB and Rewards** | **Mix FB and Rewards** |
| OVERALL TREATMENT EFFECTS | | | | | | |
| Within class social comparison | 0.084 | **0.112*** | | | 0.086 | **0.099*** |
| (Treatment 1) | (0.081) | **(0.059)** | | | (0.079) | **(0.059)** |
| Across class social comparison | 0.024 | **0.093**** | | | 0.046 | **0.089§** |
| (Treatment 2) | (0.084) | **(0.055)** | | | (0.081) | **(0.056)** |
| Financial Rewards | | | **0.231**** | **0.151*** | **0.233**** | **0.142*** |
| | | | **(0.092)** | **(0.082)** | **(0.093)** | **(0.078)** |
| Repurational Rewards | | | **0.185**** | **0.127*** | **0.184**** | **0.115*** |
| | | | **(0.079)** | **(0.066)** | **(0.078)** | **(0.064)** |
| Controlled for stratas | No | Yes | No | Yes | No | Yes |
| Interactions | No | No | No | No | No | No |
| N | 5102 | 5102 | 5102 | 5102 | 5102 | 5102 |

Note: Robust standard errors adjusted for clustering at class level are in parentheses. Controlled for stratum fixed effects (except columns (1) and (4)) - 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 15%; * significant at 10%; ** significant at 5%; *** significant at 1%

Table 7b: OLS ESTIMATES OF THE EFFECTS OF DIFFERENT MOTIVATION SCHEMES ON STUDENTS' PERFORMANCE IN ENGLISH

| **Dependent variable: Math and English score** | **ENGLISH** | | | | | |
|---|---|---|---|---|---|---|
| | **Pure FB** | **Pure FB** | **Pure Rewards** | **Pure Rewards** | **Mix FB and Rewards** | **Mix FB and Rewards** |
| OVERALL TREATMENT EFFECTS | | | | | | |
| Within class social comparison | **-0.102§** | -0.015 | | | **-0.099*** | -0.028 |
| (Treatment 1) | **(0.067)** | (0.042) | | | **(0.058)** | (0.039) |
| Across class social comparison | -0.039 | 0.014 | | | -0.007 | 0.012 |
| (Treatment 2) | (0.071) | (0.042) | | | (0.064) | (0.040) |
| Financial Rewards | | | **0.336***** | **0.153**** | **0.340***** | **0.158**** |
| | | | **(0.055)** | **(0.066)** | **(0.052)** | **(0.053)** |
| Repurational Rewards | | | **0.250**** | **0.103*** | **0.254***** | **0.108**** |
| | | | **(0.066)** | **(0.054)** | **(0.067)** | **(0.053)** |
| Controlled for stratas | No | Yes | No | Yes | No | Yes |
| Interactions | No | No | No | No | No | No |
| N | 5093 | 5093 | 5093 | 5093 | 5093 | 5093 |

Note: Robust standard errors adjusted for clustering at class level are in parentheses. Controlled for stratum fixed effects (except columns (1) and (4)) - 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 15%; * significant at 10%; ** significant at 5%; *** significant at 1%

Table 7c: OLS ESTIMATES OF THE EFFECTS OF DIFFERENT MOTIVATION SCHEMES ON STUDENTS' PERFORMANCE IN MATHEMATICS AND ENGLISH SEPARATELY

| | **Mathematics** | **English** |
|---|---|---|
| A. INTERACTION OF THE TREATMENTS | | |
| Pure within class social comparison (T1_SOLO) | 0.100 | **-0.128**** |
| | (0.084) | **(0.056)** |
| Pure across class social comparison (T2_SOLO) | 0.082 | -0.049 |
| | (0.074) | (0.059) |
| Pure inancial Rewards (Fin_SOLO) | 0.106 | 0.045 |
| | (0.101) | (0.088) |
| Pure reputational Rewards (Rep_SOLO) | 0.138 | 0.016 |
| | (0.141) | (0.082) |
| Within class comparison with financial reward (T1_fin) | **0.231*** | 0.103 |
| | **(0.118)** | (0.094) |
| Within class social comparison with reputational reward (T1_rep) | **0.209**** | 0.087 |
| | **(0.103)** | (0.080) |
| Across class social comparison with financial reward (T2_fin) | **0.277**** | **0.173*** |
| | **(0.139)** | **(0.094)** |
| Across class social comparison with reputational reward (T2_rep) | **0.188**** | 0.047 |
| | **(0.080)** | (0.080) |
| Baseline Math/English score | **0.729***** | **0.737***** |
| | **(0.017)** | **(0.016)** |
| Controlled for stratas | Yes | Yes |
| N | 5102 | 5093 |

Note: Robust standard errors adjusted for clustering at class level are in parentheses. In columns (2) and (4) controlled for stratum fixed effects - 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 15%; * significant at 10%; ** significant at 5%; *** significant at 1%

Which of the incentives dominate? Comparison of the interaction terms of the treatments with pure feedback and pure reward groups allow me to look at the value added of the combination of the treatments. The results differ by subject and by gender and can be seen in Table 8c. Girls exposed to pure feedback provision without any incentives improve by 0.12 – 0.14 standard deviations in Mathematics. The combination of the feedback together with rewards lead to girls' improvements as well, however tests reveal that there are no significant differences between the effects of pure feedback provisions with no rewards and the combination of feedback with rewards, suggesting that the improvements are driven mainly by feedback itself. Girls do not seem to react to

Table 8a: OLS ESTIMATES OF THE EFFECTS OF DIFFERENT MOTIVATION SCHEMES ON STUDENTS' PERFORMANCE IN MATHEMATICS – BY GENDER

| | MATHEMATICS | | | | | |
|---|---|---|---|---|---|---|
| **Dependent variable: Math score** | **GIRLS** | | | **BOYS** | | |
| | **Pure FB** | **Pure Rewards** | **FB and Rewards** | **Pure FB** | **Pure Rewards** | **FB and Rewards** |
| A. OVERALL EFFECTS OF TREATMENTS | | | | | | |
| Within class social comparison (Treatment 1) | **0.157*** (0.058)** | | **0.149** (0.058)** | 0.059 (0.079) | | 0.038 (0.071) |
| Across class social comparison (Treatment 2) | **0.163*** (0.060)** | | **0.159*** (0.061)** | 0.005 (0.070) | | 0.003 (0.065) |
| Financial Rewards | | 0.103 (0.096) | 0.088 (0.088) | | **0.214** (0.089)** | **0.207** (0.089)** |
| Repurational Rewards | | 0.087 (0.076) | 0.062 (0.071) | | **0.173** (0.077)** | **0.170** (0.073)** |
| Controlled for stratas | Yes | Yes | Yes | Yes | Yes | Yes |
| N | 2858 | 2858 | 2858 | 2207 | 2207 | 2207 |

Table 8b: OLS ESTIMATES OF THE EFFECTS OF DIFFERENT MOTIVATION SCHEMES ON STUDENTS' PERFORMANCE IN ENGLISH – BY GENDER

| | ENGLISH | | | | | |
|---|---|---|---|---|---|---|
| **Dependent variable: English score** | **GIRLS** | | | **BOYS** | | |
| | **Pure FB** | **Pure Rewards** | **FB and Rewards** | **Pure FB** | **Pure Rewards** | **FB and Rewards** |
| A. OVERALL EFFECTS OF TREATMENTS | | | | | | |
| Within class social comparison (Treatment 1) | -0.016 (0.045) | | -0.027 (0.042) | -0.022 (0.057) | | -0.038 (0.051) |
| Across class social comparison (Treatment 2) | 0.019 (0.048) | | 0.014 (0.045) | 0.001 (0.056) | | 0.005 (0.051) |
| Financial Rewards | | 0.089 (0.069) | 0.094 (0.068) | | **0.226*** (0.078)** | **0.234*** (0.078)** |
| Repurational Rewards | | **0.096§ (0.058)** | **0.099* (0.056)** | | **0.106§ (0.066)** | **0.111* (0.067)** |
| Controlled for stratas | Yes | Yes | Yes | Yes | Yes | Yes |
| N | 2858 | 2858 | 2858 | 2207 | 2207 | 2207 |

Note: Robust standard errors adjusted for clustering at class level are in parentheses. Controlled for stratum fixed effects - 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 15%; * significant at 10%; ** significant at 5%; *** significant at 1%

Table 8c: OLS ESTIMATES OF THE EFFECTS OF DIFFERENT MOTIVATION SCHEMES ON STUDENTS' PERFORMANCE IN MATHEMATICS AND ENGLISH SEPARATELY

| Dependent variable: Math or English score | MATHEMATICS | | ENGLISH | |
|---|---|---|---|---|
| | Girls | Boys | Girls | Boys |
| INTERACTION OF THE TREATMENTS | | | | |
| Within class social comparison (T1) | **0.121§** **(0.080)** | 0.076 (0.107) | **-0.139**** **(0.059)** | **-0.115§** **(0.072)** |
| Across class social comparison (T2) | **0.137*** **(0.077)** | 0.009 (0.088) | -0.075 (0.066) | -0.019 (0.072) |
| Financial Rewards (Fin) | 0.023 (0.103) | **0.208*** **(0.125)** | -0.034 (0.096) | 0.143 (0.113) |
| Reputational Rewards (Rep) | 0.043 (0.189) | 0.218 (0.210) | -0.062 (0.084) | 0.061 (0.123) |
| Within class comparison financial reward (T1_fin) | **0.228*** **(0.117)** | **0.225§** **(0.139)** | 0.014 (0.092) | **0.198*** **(0.116)** |
| Within class social comparison reputational reward (T1_rep) | **0.203**** **(0.101)** | **0.202§** **(0.131)** | 0.070 (0.088) | 0.094 (0.095) |
| Across class social comparison financial reward (T2_fin) | **0.278*** **(0.159)** | **0.287§** **(0.175)** | 0.114 (0.099) | **0.257**** **(0.113)** |
| Across class social comparison reputational reward (T2_rep) | **0.187**** **(0.091)** | **0.175*** **(0.104)** | 0.043 (0.083) | 0.044 (0.103) |
| Baseline Math/English score | **0.748***** **(0.021)** | **0.703***** **(0.022)** | **0.728***** **(0.019)** | 0.750*** (0.019) |
| Controlled for stratas | Yes | Yes | Yes | Yes |
| N | 2798 | 2145 | 2793 | 2142 |

Note: Robust standard errors adjusted for clustering at class level are in parentheses. Controlled for stratum fixed effects (except columns (1) and (4)) - 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%

pure rewards. Boys, on contrary, do not react to any of the feedback provisions and improve their performance mainly in response to financial rewards. In their case there is no value added if the feedback is provided on top of the rewards. However there is a significant improvement if students have received class-comparative feedback and they are offered financial rewards – those students improve by 0.274 in case of within class comparison (T1) and 0.144 in case of across class comparison (T2). Overall the results suggest that girls take into consideration how do they "look like" in their proximate circles, boys are more materialistically oriented and care about money only. Similarly to the previous results, the effect of feedback seem to fade away once their expectations

are not met and therefore we observe improvements only in Mathematics. Boys improve in response to rewards in both subjects. We may conclude that provision of feedback (considered as pure status incentives) serve as different motivation mechanisms compared to financial or reputation rewards.

Naturally, the question is whether the treatments influence all students in the same way in terms of their academic performances. In Tables 9 up to 11, I provide the results of the ordinary least squares regressions for all four quartiles based on students' baseline performances. The pattern of the effects is similar to the ones already discussed, girls of all ability quartile react mostly to feedback provision and boys of all ability quartile mostly to rewards, the effects however differ in their strengths and significance.

Table 9: OLS REGRESSIONS OF THE EFFECTS OF DIFFERENT MOTIVATION SCHEMES ON STUDENTS' PERFORMANCE IN **MATHEMATICS, BY THEIR INITIAL PERFORMANCE AND BY GENDER**

| **Dependent variable: Math score** | **GIRLS** | | | | **BOYS** | | | |
|---|---|---|---|---|---|---|---|---|
| | **Quart 1** | **Quart 2** | **Quart 3** | **Quart 4** | **Quart 1** | **Quart 2** | **Quart 3** | **Quart 4** |
| Within class social comparison (T1) | **0.126**** | 0.087 | 0.135 | **0.486***** | **-0.159**** | -0.020 | -0.006 | 0.152 |
| | **(0.060)** | (0.070) | (0.111) | **(0.106)** | **(0.074)** | (0.096) | (0.097) | (0.131) |
| Across class social comparison(T2) | 0.049 | **0.126§** | **0.242**** | **0.476***** | **-0.133*** | -0.067 | 0.026 | **0.233**** |
| | (0.054) | **(0.075)** | **(0.123)** | **(0.142)** | **(0.077)** | (0.112) | (0.114) | **(0.094)** |
| Financial Rewards | -0.057 | 0.037 | **0.185§** | 0.076 | 0.089 | **0.309**** | **0.255§** | 0.159 |
| | (0.080) | (0.090) | **(0.123)** | (0.178) | (0.095) | **(0.123)** | **(0.157)** | (0.126) |
| Repurational Rewards | 0.002 | 0.094 | -0.031 | 0.052 | **0.156*** | **0.237**** | 0.123 | 0.015 |
| | (0.082) | (0.091) | (0.123) | (0.135) | **(0.080)** | **(0.118)** | (0.135) | (0.109) |

Table 10: OLS REGRESSIONS OF THE EFFECTS OF DIFFERENT MOTIVATION SCHEMES ON STUDENTS' PERFORMANCE IN **ENGLISH**, **BY THEIR INITIAL PERFORMANCE AND BY GENDER** (continued)

| | **GIRLS** | | | | **BOYS** | | | |
|---|---|---|---|---|---|---|---|---|
| | **Quart 1** | **Quart 2** | **Quart 3** | **Quart 4** | **Quart 1** | **Quart 2** | **Quart 3** | **Quart 4** |
| Within class social comparison (T1) | -0.028 | -0.038 | 0.025 | -0.059 | 0.027 | -0.005 | -0.099 | 0.002 |
| | (0.063) | (0.066) | (0.066) | (0.086) | (0.097) | (0.086) | (0.079) | (0.083) |
| Across class social comparison(T2) | -0.001 | -0.011 | 0.032 | 0.072 | -0.005 | -0.014 | -0.034 | 0.076 |
| | (0.078) | (0.062) | (0.069) | (0.091) | (0.085) | (0.087) | (0.095) | (0.076) |
| Financial Rewards | -0.064 | 0.007 | **0.195**** | **0.301***** | 0.151 | **0.171*** | **0.357***** | **0.426***** |
| | (0.103) | (0.101) | **(0.091)** | **(0.094)** | (0.122) | **(0.098)** | **(0.099)** | **(0.096)** |
| Repurational Rewards | -0.006 | 0.024 | **0.211**** | 0.101 | -0.008 | **0.146*** | **0.180**** | 0.121 |
| | (0.106) | (0.080) | **(0.096)** | (0.082) | (0.127) | **(0.084)** | **(0.089)** | (0.097) |

Note: Robust standard errors adjusted for clustering at class level are in parentheses. Controlled for stratum fixed effects - 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%

Table 11a: OLS REGRESSIONS OF THE EFFECTS OF DIFFERENT MOTIVATION SCHEMES ON STUDENTS' PERFORMANCE IN **MATHEMATICS, BY THEIR INITIAL PERFORMANCE AND BY GENDER**

| **Dependent variable: Math score** | **GIRLS** | | | | **BOYS** | | | |
|---|---|---|---|---|---|---|---|---|
| | **Quart 1** | **Quart 2** | **Quart 3** | **Quart 4** | **Quart 1** | **Quart 2** | **Quart 3** | **Quart 4** |
| Within class social comparison | 0.098 | 0.019 | 0.139 | 0.188 | -0.097 | -0.073 | 0.014 | **0.354**** |
| (T1_solo) | (0.087) | (0.102) | (0.143) | (0.133) | (0.090) | (0.116) | (0.205) | **(0.147)** |
| Across class social comparison | **0.195*** | 0.070 | (0.217)* | 0.088 | -0.108 | -0.146 | 0.085 | 0.139 |
| (T2_solo) | **(0.099)** | (0.108) | (0.130) | (0.136) | (0.095) | (0.122) | (0.156) | (0.108) |
| Financial Rewards (fin_solo) | 0.011 | -0.093 | 0.091 | -0.234 | 0.180 | **0.257§** | 0.254 | 0.090 |
| | (0.095) | (0.144) | (0.183) | (0.233) | (0.149) | **(0.168)** | (0.218) | (0.186) |
| Repurational Rewards | 0.052 | 0.100 | 0.043 | **-0.323§** | 0.150 | 0.156 | 0.204 | 0.192 |
| (rep_solo) | (0.119) | (0.158) | (0.271) | **(0.216)** | (0.150) | (0.232) | (0.198) | (0.203) |
| Within class comparison | 0.139 | **0.189§** | 0.272 | **0.270§** | -0.082 | 0.146 | 0.211 | **0.363§** |
| financial reward (T1_fin) | (0.126) | **(0.129)** | (0.197) | **(0.186)** | (0.121) | (0.155) | (0.208) | **(0.234)** |
| Within class comparison | **0.203*** | 0.060 | 0.128 | **0.444***** | 0.034 | **0.281*** | 0.214 | 0.022 |
| reputational reward (T1_rep) | **(0.121)** | (0.110) | (0.182) | **(0.165)** | (0.115) | **(0.167)** | (0.224) | (0.199) |
| Across class comparison | -0.028 | 0.103 | **0.549**** | **0.469§** | -0.037 | 0.305 | 0.400 | **0.518***** |
| financial reward (T2_fin) | (0.107) | (0.145) | **(0.229)** | **(0.314)** | (0.134) | (0.258) | (0.317) | **(0.157)** |
| Across class comparison | 0.043 | 0.212 | 0.094 | **0.313**** | 0.078 | 0.055 | 0.051 | **0.325**** |
| reputational reward (T2_rep) | (0.092) | (0.149) | (0.173) | **(0.128)** | (0.111) | (0.156) | (0.212) | **(0.125)** |
| Baseline Math/English score | **0.509***** | **0.916***** | **0.925***** | **0.689***** | **0.499***** | **0.890***** | **0.938***** | **0.676***** |
| | **(0.124)** | **(0.141)** | **(0.159)** | **(0.045)** | **(0.112)** | **(0.188)** | **(0.166)** | **(0.044)** |
| Controlled for stratas | Yes | Yes | Yes | Yes | Yes | Yes | Yes | Yes |
| N | 760 | 794 | 728 | 576 | 443 | 602 | 560 | 602 |

Table 11b: OLS REGRESSIONS OF THE EFFECTS OF DIFFERENT MOTIVATION SCHEMES ON STUDENTS' PERFORMANCE IN **ENGLISH**, **BY THEIR INITIAL PERFORMANCE AND BY GENDER**

| **Dependent variable: English score** | **GIRLS** | | | | **BOYS** | | | |
|---|---|---|---|---|---|---|---|---|
| | **Quart 1** | **Quart 2** | **Quart 3** | **Quart 4** | **Quart 1** | **Quart 2** | **Quart 3** | **Quart 4** |
| Within class social comparison (T1_solo) | **-0.208*** **(0.114)** | **-0.205**** **(0.089)** | **-0.175*** **(0.095)** | -0.052 (0.076) | **-0.278*** **(0.151)** | -0.046 (0.088) | **-0.223§** **(0.136)** | 0.077 (0.127) |
| Across class social comparison (T2_solo) | -0.094 (0.131) | **-0.241***** **(0.088)** | **-0.191*** **(0.115)** | 0.011 (0.070) | -0.136 (0.163) | -0.039 (0.098) | -0.127 (0.122) | -0.001 (0.091) |
| Financial Rewards (fin_solo) | **-0.272*** **(0.164)** | **-0.216*** **(0.124)** | 0.001 (0.126) | **0.302§** **0.187** | -0.087 (0.226) | 0.031 (0.166) | **0.287*** **(0.153)** | **0.343***** **(0.111)** |
| Repurational Rewards (rep_solo) | -0.072 (0.176) | -0.150 (0.107) | -0.018 (0.161) | 0.045 (0.176) | -0.206 (0.185) | **0.221§** **(0.135)** | 0.032 (0.155) | 0.202 (0.191) |
| Within class comparison financial reward (T1_fin) | -0.065 (0.170) | -0.127 (0.169) | 0.041 (0.128) | **0.158*** **(0.082)** | 0.137 (0.244) | **0.241§** **(0.154)** | **0.199§** **(0.135)** | **0.287***** **(0.093)** |
| Within class comparison reputational reward (T1_rep) | -0.155 (0.173) | -0.149 (0.121) | **0.193§** **(0.124)** | 0.064 (0.086) | -0.071 (0.232) | 0.041 (0.111) | 0.046 (0.122) | **0.187§** **(0.114)** |
| Across class comparison financial reward (T2_fin) | -0.158 (0.196) | -0.084 (0.145) | 0.178 (0.124) | **0.403***** **(0.087)** | -0.011 (0.188) | 0.157 (0.144) | 0.232 (0.163) | **0.724***** **(0.105)** |
| Across class comparison reputational reward (T2_rep) | -0.094 (0.194) | -0.075 (0.119) | 0.093 (0.136) | **0.152**** **(0.070)** | -0.160 (0.233) | 0.092 (0.133) | 0.113 (0.169) | 0.049 (0.094) |
| Baseline Math/English score | **0.871***** **(0.129)** | **0.829***** **(0.118)** | **0.708***** **(0.092)** | **0.607***** **(0.043)** | **1.037***** **(0.126)** | **0.729***** **(0.123)** | **0.735***** **(0.115)** | **0.540***** **(0.043)** |
| Controlled for stratas | Yes | Yes | Yes | Yes | Yes | Yes | Yes | Yes |
| N | 689 | 746 | 693 | 726 | 572 | 570 | 496 | 564 |

### 4e. The effect of treatments on attrition

Estimates of treatment effects can be biased if the attrition from control versus treatment groups systematically differs and the difference is caused by the presence of the treatment. As shown in Table 3 and Appendix C, students in treatment groups attrite less often in absolute values and are more often present in all five testing rounds compared to their control-group counterparts. In order to see whether and to what extent social comparison and reward treatments influence probability to dropout, I run probit model on attrition and full attendance on all treatments controlling for strata variables (Table 12).

Table 12: TREATMENT EFFECTS ON PROBABILITIES OF STUDENTS' ATTENDANCES

| Overall treatment effects on: | **Attrition** | **Alwayscomer** |
|---|---|---|
| Within class social comparison (T1) | **-0.088*** (0.028)** | **0.120*** (0.035)** |
| Across class social comparison (T2) | **-0.111*** (0.026)** | **0.108*** (0.032)** |
| Financial Rewards (Fin) | **-0.122*** (0.029)** | **0.124*** (0.038)** |
| Reputational Rewards (Rep) | **-0.126*** (0.035)** | 0.034 (0.043) |
| Controlled for stratas | Yes | Yes |
| N | 7109 | 7109 |

Note: Robust standard errors adjusted for clustering at class level are in parentheses. Controlled for gender and stratum fixed effects - 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%

Attrition rate comprises of students who missed our last testing round but attended the baseline testing at the beginning of the project. Non-rewarded students exposed to both within and across class social comparison feedback have from 8.8 to 11 per cent lower probability to miss the final testing round. Similarly, rewarded students without feedback have by approximately 12 per cent lower probability to attrite. As discussed in section 4c, despite the different attrition across

treatment and control groups, students who remained at schools in the last testing round are on average the same in terms of initial characteristics and therefore the OLS estimates should not be biased. In the following section I run different specifications to compare OLS estimates with estimates that correct for possible attrition bias.

### 4f. Stability of the results

In order to adjust the results for non-random attrition, I proceeded with imputation methods and inverse probability-weighted regressions (Imbens, 2004; Woolridge, 2007; Kwak (2010), Hirano et al., 2000, etc.). Inverse probability weighting (IPW) can adjust for confounding factors and selection bias. As the title suggests, IPW assigns a weight to every student which equals to the student's inverse probability to be absent/to attrite and adjust for that in estimation of the treatment effects. Imputation method is used to fill the missing observations of students who were absent or dropout in the last testing round based on a predefined rule.

Table 13 provides the comparison of ordinary least squares estimations (columns 1 and 2) of the treatment effects to the weighted least squares using inverse probability weights (columns 3 to 4), separately for Math and English. Correcting for the probability to dropout, treatment effects are similar or slightly higher in absolute terms but not significantly different.

The results of the imputation methods (shown in Table 13) bring similar conclusions. I use three different measures to impute missing observations – median ratio, overall percentile ranks and the class percentile ranks (inspired by Krueger, 1999). All of the measures take the advantage of repeated school visits and follow the same logic – if the observation from the last school visit is missing, I look at the last score available and adjust for the differences in test difficulty. The same procedure is done to impute Math and English scores separately. Median ratio measure imputes the last available observation and adjusts it by class median in the last round divided by the median in the last available round. Overall percentile ranks and the class percentile ranks take into

consideration the rank of the student in the last available distribution and impute the score corresponding to the student of the same rank in the final visit distribution. The two measures differ in the group within which the rank order is calculated. While in the overall percentile rank, I look at the student's rank order within respective treatment 1, treatment 2 and control groups, in the class percentile ranks, I assign the rank based on the student's particular class rank position. Imputation method artificially fills missing observations and the results serve only as bounds.

All three imputation measures deliver similar results to initial ordinary least squares in terms of the significance of the treatment effect and the signs of the effects. Median ratio and class percentile ranks are in line with the downward bias of the average treatment effects computed by ordinary least squares regression and are comparable to the weighted regression estimates. Ordinary least squares estimates based on the overall percentile rank imputation are comparable.

Table 13a: COMPARISON OF THE ESTIMATES OF THE EFFECTS OF DIFFERENT MOTIVATION SCHEMES ON STUDENTS' PERFORMANCE IN MATHEMATICS

| | MATHEMATICS | | | | |
|---|---|---|---|---|---|
| **Dependent variable: Math and English score** | **OLS** | **IPW** | **Imputation (median ratio)** | **Imputation (overall percentiles)** | **Imputation (class percentiles)** |
| OVERALL EFFECTS OF TREATMENTS | | | | | |
| Within class social comparison | **0.099*** | 0.080 | **0.101§** | **0.093*** | **0.094*** |
| (Treatment 1) | **(0.059)** | (0.066) | **(0.067)** | **(0.051)** | **(0.053)** |
| Across class social comparison | **0.089§** | **0.125*** | **0.090§** | **0.086*** | **0.082§** |
| (Treatment 2) | **(0.056)** | **(0.066)** | **(0.059)** | **(0.051)** | **(0.055)** |
| Financial Rewards | **0.142*** | **0.224**** | **0.173**** | 0.093 | **0.125*** |
| | **(0.078)** | **(0.087)** | **(0.075)** | (0.068) | **(0.074)** |
| Repurational Rewards | **0.115*** | **0.133*** | **0.132**** | **0.092*** | **0.107*** |
| | **(0.064)** | **(0.079)** | **(0.066)** | **(0.055)** | **(0.059)** |
| Controlled for stratas | Yes | Yes | Yes | Yes | Yes |
| N | 5102 | 5102 | 6736 | 7107 | 7041 |

Table 13b: COMPARISON OF THE ESTIMATES OF THE EFFECTS OF DIFFERENT MOTIVATION SCHEMES ON STUDENTS' PERFORMANCE IN ENGLISH

| | ENGLISH | | | | |
|---|---|---|---|---|---|
| **Dependent variable: Math and English score** | **OLS** | **IPW** | **Imputation (median ratio)** | **Imputation (overall percentiles)** | **Imputation (class percentiles)** |
| OVERALL EFFECTS OF TREATMENTS | | | | | |
| Within class social comparison | -0.028 | -0.004 | 0.043 | -0.024 | -0.009 |
| (Treatment 1) | (0.039) | (0.044) | (0.053) | (0.034) | (0.042) |
| Across class social comparison | 0.012 | **0.069§** | 0.060 | 0.024 | 0.009 |
| (Treatment 2) | (0.040) | **(0.044)** | (0.051) | (0.035) | (0.044) |
| Financial Rewards | **0.158**** | **0.190***** | **0.129*** | **0.144***** | **0.135**** |
| | **(0.053)** | **(0.063)** | **(0.068)** | **(0.052)** | **(0.064)** |
| Repurational Rewards | **0.108**** | **0.109*** | 0.063 | **0.116***** | 0.065 |
| | **(0.053)** | **(0.057)** | (0.064) | **(0.040)** | (0.053) |
| Controlled for stratas | Yes | Yes | Yes | Yes | Yes |
| N | 5093 | 5093 | 6736 | 7107 | 7040 |

Note: Robust standard errors adjusted for clustering at class level are in parentheses. Controlled for stratum fixed effects (except columns (1) and (4)) - 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 15%; * significant at 10%; ** significant at 5%; *** significant at 1%

Table 14a: COMPARISON OF THE ESTIMATES OF THE EFFECTS OF DIFFERENT MOTIVATION SCHEMES ON STUDENTS' PERFORMANCE IN MATHEMATICS

| | MATHEMATICS | | | | |
|---|---|---|---|---|---|
| **Dependent variable: Math or English score** | **OLS** | **IPW** | **Imputation (median ratio)** | **Imputation (overall percentiles)** | **Imputation (class percentiles)** |
| INTERACTION OF THE TREATMENTS | | | | | |
| Within class social comparison (T1) | 0.100 | 0.035 | **0.133*** | 0.064 | 0.070 |
| | (0.084) | (0.091) | **(0.079)** | (0.073) | (0.082) |
| Across class social comparison (T2) | 0.082 | 0.061 | **0.129*** | 0.020 | 0.036 |
| | (0.074) | (0.081) | **(0.068)** | (0.058) | (0.074) |
| Financial Rewards (Fin) | 0.106 | 0.112 | **0.169*** | 0.004 | 0.070 |
| | (0.101) | (0.099) | **(0.096)** | (0.079) | (0.097) |
| Reputational Rewards (Rep) | 0.138 | 0.135 | **0.206*** | 0.085 | 0.092 |
| | (0.141) | (0.136) | **(0.124)** | (0.105) | (0.115) |
| Within class comparison financial reward (T1_fin) | **0.231*** | **0.267**** | **0.281**** | **0.170§** | **0.202*** |
| | **(0.118)** | **(0.132)** | **(0.129)** | **(0.103)** | **(0.116)** |
| Within class social comparison reputational reward (T1_rep) | **0.209**** | **0.187§** | **0.266**** | **0.134§** | **0.171*** |
| | **(0.103)** | **(0.114)** | **(0.112)** | **(0.088)** | **(0.098)** |
| Across class social comparison financial reward (T2_fin) | **0.277**** | **0.389***** | **0.331**** | **0.188§** | **0.209§** |
| | **(0.139)** | **(0.136)** | **(0.128)** | **(0.115)** | **(0.130)** |
| Across class social comparison reputational reward (T2_rep) | **0.188**** | **0.174*** | **0.186**** | **0.139**** | **0.164**** |
| | **(0.080)** | **(0.089)** | **(0.073)** | **(0.063)** | **(0.076)** |
| Baseline Math/English score | **0.729***** | **0.746***** | **0.755***** | **0.679***** | **0.658***** |
| | **(0.017)** | **(0.019)** | **(0.048)** | **(0.016)** | (0.019) |
| Controlled for stratas | Yes | Yes | Yes | Yes | Yes |
| N | 5102 | 5102 | 6736 | 7107 | 7041 |

Note: Robust standard errors adjusted for clustering at class level are in parentheses. In columns (2) and (4) controlled for stratum fixed effects - 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 15%; * significant at 10%; ** significant at 5%; *** significant at 1%

Table 14b: COMPARISON OF THE ESTIMATES OF THE EFFECTS OF DIFFERENT MOTIVATION SCHEMES ON STUDENTS' PERFORMANCE IN ENGLISH

| | ENGLISH | | | | |
|---|---|---|---|---|---|
| **Dependent variable: Math or English score** | **OLS** | **IPW** | **Imputation (median ratio)** | **Imputation (overall percentiles)** | **Imputation (class percentiles)** |
| INTERACTION OF THE TREATMENTS | | | | | |
| Within class social comparison (T1) | **-0.128**** **(0.056)** | **-0.134*** **(0.069)** | **-0.133**** **(0.060)** | **-0.135***** **(0.045)** | **-0.207***** **(0.062)** |
| Across class social comparison (T2) | -0.049 (0.059) | -0.087 (0.070) | -0.052 (0.063) | -0.046 (0.048) | **-0.139**** **(0.065)** |
| Financial Rewards (Fin) | 0.045 (0.088) | 0.021 (0.084) | -0.006 (0.096) | 0.041 (0.069) | -0.047 (0.093) |
| Reputational Rewards (Rep) | 0.016 (0.082) | -0.008 (0.085) | -0.089 (0.123) | 0.036 (0.059) | -0.099 (0.086) |
| Within class comparison financial reward (T1_fin) | 0.103 (0.094) | 0.112 (0.086) | 0.096 (0.108) | 0.072 (0.080) | 0.043 (0.101) |
| Within class social comparison reputational reward (T1_rep) | 0.087 (0.080) | 0.022 (0.085) | 0.069 (0.082) | 0.069 (0.058) | -0.009 (0.081) |
| Across class social comparison financial reward (T2_fin) | **0.173*** **(0.094)** | **0.216**** **(0.104)** | 0.113 (0.099) | **0.137*** **(0.075)** | 0.062 (0.104) |
| Across class social comparison reputational reward (T2_rep) | 0.047 (0.080) | 0.093 (0.079) | 0.024 (0.082) | 0.059 (0.064) | -0.042 (0.087) |
| Baseline Math/English score | **0.737***** **(0.016)** | **0.698***** **(0.021)** | **0.738***** **(0.025)** | **0.701***** **(0.014)** | **0.682***** **(0.017)** |
| Controlled for stratas | Yes | Yes | Yes | Yes | Yes |
| N | 5093 | 5093 | 6736 | 7017 | 7040 |

Note: Robust standard errors adjusted for clustering at class level are in parentheses. In columns (2) and (4) controlled for stratum fixed effects - 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 15%; * significant at 10%; ** significant at 5%; *** significant at 1%

## 5. Conclusion

This paper contributes to the current literature by studying the effects of social comparison on students' performance. It directly allows differentiating between within and across class social comparisons. 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 Bandiera (2011) found negative effects. In my intervention, feedback provision seems to have small but positive overall effect of the size 0.1 standard deviation. The overall effect is pushed down by the heterogeneity of the effect across gender. Once it is decomposed one can see that while girls improved by 0.16 standard deviations, boys performance remained the same in response to feedback provision. Two types of rewards were introduced orthogonally to the feedback treatments – financial rewards (2000 Ugandan Shillings) and reputational rewards (winners' names announced in local newspapers). The rewards (on contrary to feedback) motivate boys to improve their performance (2.34 to 2.7 standard deviation in Math and 1.1 to 1.7 standard deviations in English). Both girls and boys react similarly to the interaction of feedback and rewards, however their driving mechanisms differ. Girls seem to care more about their status or reputation among the proximate classmates and the main influence come from feedback provision (there is no significant value added of rewards to students' performance). Boys performance is driven by the provision of rewards as there is insignificant value added of feedback once rewards are provided. Both social comparison treatments have an effect on students' attrition rates (treatment group have by 9 to 13 per cent lower attrition). I run different specifications to compare OLS estimates with estimates that correct for possible attrition bias. The adjusted results suggest that ordinary least squares estimates are similar to estimates of different specifications.

**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.

Hirano, K., Imbens, G. and G. Ridder (2003): Efficient Estimation of Average Treatment Effects Using the Estimated Propensity Score, Econometrica, Vol. 71(4), 1161–1189

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

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, A. B. (1999): Experimental estimates of education production functions, The Quarterly Journal of Economics, 114 (2): 497-532

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

Wooldridge, J. (2007): Inverse Probability Weighted M-Estimation for General Missing Data Problems, Journal of Econometrics 141:1281-1301

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 | 2415 | 2371 |
| | (48 groups) | (51 groups) | (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: MEAN 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: COMPARISON OF MEAN CHARACTERISTICS OF STUDENTS' IN TREATMENT AND CONTROL GROUPS (Continued)

| | Means | | | Mean Differences | | Joint P-value |
|---|---|---|---|---|---|---|
| | T1 | T2 | Control | (T1 – C) | (T2 – C) | |
| C. OTHER (continued) | | | | | | |
| **D.1 Attrition rates** | | | | | | |
| All schools | 0.359 | 0.346 | 0.454 | **-0.095*** ** | **-0.108*** ** | 0.002 |
| | | | | **(0.034)** | **(0.033)** | |
| Restricted sample# | 0.358 | 0.348 | 0.417 | **-0.059*** | **-0.069** ** | 0.041 |
| | | | | **(0.030)** | **(0.029)** | |
| **D.2 Dropouts** | | | | | | |
| All schools | 0.094 | 0.088 | 0.139 | **-0.044*** ** | **-0.051*** ** | 0.009 |
| | | | | **(0.016)** | **(0.017)** | |
| Restricted sample# | 0.091 | 0.087 | 0.113 | **-0.022*** | **-0.025** ** | 0.158 |
| | | | | **(0.013)** | **(0.014)** | |
| **D.3 Absences** | | | | | | |
| All schools | 0.045 | 0.044 | 0.055 | -0.009 | **-0.011*** | 0.222 |
| | | | | (0.008) | **(0.007)** | |
| Restricted sample# | 0.045 | 0.044 | 0.056 | -0.011 | **-0.013*** | 0.222 |
| | | | | (0.008) | **(0.007)** | |
| **D.4 Comers** | | | | | | |
| All schools | 0.529 | 0.565 | 0.520 | 0.009 | **0.045*** | 0.165 |
| | | | | (0.019) | **(0.024)** | |
| Restricted sample# | 0.543 | 0.575 | 0.549 | -0.007 | 0.026 | 0.395 |
| | | | | (0.020) | (0.024) | |
| **D.5 Alwayscomers** | | | | | | |
| All schools | 0.202 | 0.186 | 0.082 | **0.121*** ** | **0.104*** ** | 0.000 |
| | | | | **(0.033)** | **(0.104)** | |
| Restricted sample# | 0.207 | 0.188 | 0.110 | **0.097*** ** | **0.077** ** | 0.008 |
| | | | | **(0.033)** | **(0.031)** | |
| **D.6 Speculate** | | | | | | |
| All schools | 0.118 | 0.100 | 0.195 | **-0.077*** ** | **-0.095*** ** | 0.000 |
| | | | | **(0.025)** | **(0.023)** | |
| Restricted sample# | 0.104 | 0.089 | 0.159 | **-0.056** ** | **-0.071*** ** | 0.011 |
| | | | | **(0.026)** | **(0.024)** | |

Appendix D: OLS ESTIMATES OF THE EFFECTS OF DIFFERENT MOTIVATION SCHEMES ON STUDENTS' PERFORMANCE IN MATHEMATICS AND ENGLISH, **BY SCHOOL TYPE**

| **Dependent variable: Math and English score** | **Mathematics** | | | **English** | | |
|---|---|---|---|---|---|---|
| | **Public** | **Private** | **Community** | **Public** | **Private** | **Community** |
| A. OVERALL TREATMENT EFFECTS (LINEAR COMBINATIONS OF THE EFFECTS) | | | | | | |
| Within class social comparison | -0.022 | 0.250 | **0.634**** | -0.125 | **0.424***** | 0.100 |
| (Treatment 1) | (0.451) | (0.234) | **(0.330)** | (0.262) | **(0.104)** | (0.186) |
| Across class social comparison | 0.155 | 0.186 | 0.271 | 0.123 | **0.257**** | 0.330 |
| (Treatment 2) | (0.404) | (0.176) | (0.229) | (0.255) | **(0.104)** | 0.198 |
| Financial Rewards | **0.527*** | -0.060 | **0.772***** | 0.214 | **0.648***** | **0.644*** |
| | **(0.304)** | (0.223) | **(0.239)** | (0.274) | **(0.104)** | **(0.358)** |
| Repurational Rewards | -0.134 | 0.133 | **0.471**** | -0.100 | **0.354***** | **0.694***** |
| | (0.422) | (0.273) | **(0.174)** | (0.241) | **(0.109)** | **(0.163)** |
| B. INTERACTION OF ALL TREATMENTS | | | | | | |
| Within class social comparison | 0.104 | 0.152 | -0.017 | -0.153 | **-0.163***** | -0.129 |
| (T1) | (0.132) | (0.151) | (0.229) | (0.109) | **(0.047)** | (0.236) |
| Across class social comparison | 0.092 | 0.117 | -0.162 | -0.147 | -0.015 | 0.007 |
| (T2) | (0.151) | (0.094) | (0.240) | (0.118) | (0.055) | (0.234) |
| Financial Rewards (Fin) | 0.136 | 0.073 | 0.159 | -0.001 | 0.059 | 0.091 |
| | (0.202) | (0.158) | (0.177) | (0.183) | (0.088) | (0.254) |
| Reputational Rewards (Rep) | 0.320 | -0.168 | NA | -0.184 | 0.085 | **0.694***** |
| | (0.315) | (0.182) | | (0.154) | (0.068) | **(0.163)** |
| Within class comparison financial reward (T1_fin) | 0.135 | -0.064 | 0.179 | 0.168 | **0.347***** | 0.229 |
| | (0.229) | (0.192) | (0.217) | (0.219) | **(0.099)** | (0.285) |
| Across class comparison financial reward (T2_fin) | 0.255 | -0.070 | **0.434**** | 0.047 | **0.242***** | 0.323 |
| | (0.233) | (0.148) | **(0.195)** | (0.214) | **(0.084)** | (0.291) |
| Within class comparison reputational reward (T1_rep) | -0.262 | 0.162 | **0.471**** | -0.139 | **0.240***** | NA |
| | (0.385) | (0.243) | **(0.174)** | (0.187) | **(0.074)** | |
| Across class comparison reputational reward (T2_rep) | -0.192 | 0.139 | NA | 0.223 | 0.029 | NA |
| | (0.329) | (0.191) | | (0.186) | (0.090) | |
| Baseline Math/English score | **0.669***** | **0.723***** | **0.699***** | **0.738***** | **0.713***** | **0.756***** |
| | **(0.032)** | **(0.022)** | **(0.041)** | **(0.034)** | **(0.021)** | **(0.034)** |
| Gender | **-0.086*** | **-0.079***** | -0.019 | **0.066*** | 0.045 | **0.072*** |
| | **(0.045)** | **(0.029)** | (0.038) | **(0.035)** | (0.029) | **(0.039)** |
| Controlled for stratas | Yes | Yes | Yes | Yes | Yes | Yes |
| N | 1478 | 2532 | 1055 | 1478 | 2523 | 1055 |

Note: Robust standard errors adjusted for clustering at class level are in parentheses. Controlled for stratum fixed effects (except columns (1) and (4)) - 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%