When Student Incentives Don't Work: Evidence from a Field Experiment in Malawi

James Berry, Hyuncheol Kim, and Hyuk Son*

June 2017

PRELIMINARY AND INCOMPLETE: PLEASE DO NOT CITE OR CIRCULATE

Abstract

We study the impacts of two alternative merit-based scholarship designs for 5th to 8th graders through a field experiment in Malawi. For those in the *Standard* scholarship program, top performers on a final exam are awarded scholarships, regardless of baseline test score. For those in the *Relative* program, students are groupbed by baseline score, and scholarships are awarded to the top performers in each group. In addition, feedback in the form of students' mid-term exam scores were randomly provided to a subset of students. We find that the *Standard* program significantly *decreased* test scores and motivation to study, especially for those least likely to win the scholarship. We also find that feedback on ranking may improve test scores for high performers.

JEL Classifications: I21, O15

Keywords: student incentives, education policy, merit-based scholarship, feedback, field experiments

*Berry: Cornell University, jimberry@cornell.edu; Kim: Cornell University, hk788@cornell.edu; Son: Cornell University, hyukson@gmail.com.

1 Introduction

In recent years, performance-based incentives for students have received increasing research attention as a means to improve learning outcomes in developing countries. The results are largely mixed (Kremer, Miguel, and Thornton (2009); Sharma (2010); Li et al. (2014)).[1] However, less is known about how the structure of incentive schemes can influence performance and potentially mediate the effectiveness of these programs. The precise structure of the incentive schemes can both influence the students' performance on average, in addition to influencing the distribution of performance (e.g., Berry (2015)).

One particular incentive scheme that has received substantial attention in the literature is a standard individual tournament in which the top students on an exam are provided with a reward. Such schemes allow for the policy maker to set a fixed budget for the incentives, and have been generally shown to be incentive compatible to induce effort (Lazear and Rosen (1981)). However, tournament schemes in which relatively few students receive the reward may result in low effort from students with low levels of initial academic performance. For example, one criticism of merit-based scholarship programs is that by providing rewards to only the very top performers, lower-performing students who are unlikely to receive the incentive may not respond to the programs.

A further issue arises when students have inaccurate beliefs of their initial learning level or ranking. When a student's optimal effort varies by initial level, it is the perception of that level rather than the actual level. In many developing countries, students and parents may lack information on a student's performance in school (Dizon-Ross (2016)), and thus providing information on performance may influence align effort with initial learning levels. In addition, there is potential complementarity between information on performance and incentive in a sense that students are more likely to respond to information in performance-based incentives setting.

In this paper, we study impacts of two types of merit-based scholarships as well as feedback on academic performance of 5th to 8th graders at the primary school in Malawi. Specifically, we conduct a randomized experiment comparing the impacts of a standard tournament incentive scheme with an alternative scheme that provided rewards to students based on performance relative to a comparison group of students with similar baseline test scores. Under the relative incentive scheme,

[1] Kremer, Miguel, and Thornton (2009)find the merit-based scholarship improve test score by 0.19 standard deviation even for the low-performing female students. Li et al. (2014) find that a financial incentives for a group with high and low performers improve low performers test score by 0.265 standard deviations. However, an incentive only for low performers were ineffective. Sharma (2010) finds an increase of test score by 0.09 standard deviations on average when piece rate financial incentives on students' testing outcomes were offered, but the effects are only through high performers.

students were grouped into bins by baseline test score, and the top students within each bin received the incentive. Because students compete only with others that have similar baseline test scores, initially low-performing students may be more likely to receive the rewards compared with a standard tournament. This scheme may therefore increase effort and reduce discouragement that may accompany the standard tournament. In addition, like the standard tournament, it allows for a fixed incentive budget, as the number of students who obtain the incentive is known ex ante. The design was based on Barlevy and Neal (2012) who propose a similar scheme for teachers, which they call "pay for percentile".

The experiment was conducted in 117 classrooms in 31 primary schools in rural schools near Lilongwe, Malawi. The two incentive schemes provided a scholarship with a value of MWK 4500 (USD 9.70). Under the standard tournament scheme, was provided to students in the sample who scored in the top 15 percent on the final exam. Under the relative scheme, students were sorted into groups of 100 by baseline exam score, and the top 15 scorers on the final exam were provided with the scholarship. The study was conducted between February and June of 2015.

We find that the *Standard* scholarship scheme significantly reduced final exam scores by 8.54 standard deviations across the full sample, with the largest negative impacts on students with the lowest initial test scores. However, the *Relative* merit-based scholarship scheme did not have significant impacts on test score performance. Turning to intermediate outcomes, we find that the *Standard* scholarship scheme reduced survey-based motivation of the students, again with the results concentrated among the initially lowest-performing students. We find no significant differences in motivation for the *Relative* scholarship group relative to the control group.

There are several key implications of our results. First, standard tournament-based incentive schemes may not only have distributional implications, but they may lower performance on average. Second, schemes based on relative performance still may not increase performance if they fail to motivate students, which aligns with argument of Gneezy, Meier, and Rey-Biel (2011) that financial incentive may exclude internal motivation.

This paper contributes to the existing literature in several dimensions. First, our study is first direct comparison that we know of that compares tournament incentives with relative incentives, with a focus on impacts across the test score distribution. Our benchmark study is Kremer, Miguel, and Thornton (2009) which evaluates the effect of a merit-based girls' scholarship program in rural Kenya. The top 15% of 6th grade female students in the program districts received the scholarship for two years. They find that the merit scholarship improved average test scores by 0.19 standard deviation. However, the fact that they only have significant results in one district and do not have education impacts in the other district requires further research in a merit scholarship in developing

countries. Although they show the improvement of test score among low-scoring girls who are less likely to receive the scholarship, further heterogeneity analysis for low performing students would be needed since a structure of the merit scholarship is 'one incentive fits all' which does not directly consider weaker students with low education achievement. Blimpo (2014) conducts a randomized controlled trial of an incentive program in Benin finds impacts from 0.27 to 0.34 standard deviations for individual, team, and team tournament incentives. He also find that the tournament incentives did not affect low performing students.

This paper contributes to the existing literature in several dimensions. First, our study is first direct comparison that we know of that compares tournament incentives with relative incentives, with a focus on impacts across the test score distribution. Our benchmark study is Kremer, Miguel, and Thornton (2009) which evaluates the effect of a *Standard* merit-based girls' scholarship program in rural Kenya. The top 15% of 6th grade female students in the program districts received the scholarship for two years. They find that the *Standard* merit scholarship improved average test scores by 0.19 standard deviation. However, the fact that they only have significant results in one district and do not have education impacts in the other district requires further research in a *Standard* merit scholarship in developing countries. Although they show the improvement of test score among low-scoring girls who are less likely to receive the scholarship, further heterogeneity analysis for low performing students would be needed since a structure of the *Standard* merit scholarship is 'one incentive fits all' which does not directly consider weaker students with low education achievement.

More generally, the study contributes to the literature evaluating incentives-to-learn programs. Evidence on these programs is generally mixed, both in developing countries (Kremer, Miguel, and Thornton (2009); Sharma (2010); Li et al. (2014)) as well as in developed countries (Gneezy, Meier, and Rey-Biel (2011)).[2] However, a number of studies have indicated that the structure of incentives-to-learn programs can influence the results (Hirshleifer (2015); Behrman et al. (2015); Berry (2015); Blimpo (2014)).

Our paper is also the the first test of Barlevy and Neal (2012)'s "pay for percentile" scheme on students. Loyalka et al. (2016) evaluate this incentive structure for teachers in the context of China and find larger effects on child learning relative to absolute performance targets. The scheme is closely related to schemes that provide incentives based on improvement relative to baseline

[2]Closely related are conditional cash transfer programs which provide incentives for enrollment and attendance in school. Behrman, Sengupta, and Todd (2005) reported that participation in the PROGRESA is associated with not only improved school enrollment but also less grade repetition, lower dropout rates, and higher school reentry rates among dropouts. Baird, McIntosh, and Özler (2011) who used an experimental design to estimate the effects of conditional cash transfers in Malawi find that school enrollment and attendance as well as test scores were significantly higher among treated girls.

(Behrman et al. (2015); Berry (2015)).

Lastly, our study is related to impact of rank information on academic performance. Bandiera, Larcinese, and Rasul (2015): providing feedback on prior test scores increases subsequent exam performance. Feedback may be motivating through a sense of competition: Tran and Zeckhauser (2012), Azmat and Iriberri (2010): providing feedback on relative rank improves academic performance.

The remainder of this paper is organized as follows. Section 2 provides a description of the context and scholarship schemes. Section 3 presents the experimental design, and Section 4 presents the results. We discuss the results and conclude in Section 5.

2 Context and Programs

2.1 Primary education in Malawi

We study incentives and feedback to promote student achievement at primary schools in Malawi. Education system in Malawi is composed of eight years of primary education followed by four years of secondary education. Similar to other countries in Sub-Saharan Africa, the government of Malawi abolished primary school fees in 1994, leading to near-universal primary enrollment. However, high dropout rates leads to a low survival rate: a primary school completion is only 46% (Ministry of Education Science and Technology, 2008). In addition, enrollment drops sharply between 8th grade and secondary school, where fees are still common, to result in only 58% of secondary school transition rate (Ministry of Education Science and Technology, 2008). In addition, like many countries in the developing world, learning outcomes among Malawian primary students are low. Among the 15 countries in Sub-Saharan Africa taking the SACMEQ[3] assessments, 6th graders in Malawi scored near the bottom in both reading and mathematics. Schools are characterized by high pupil:teacher ratios, and low levels of infrastructure.[4]

Academic calendar year, starting on September, consists of three semesters. Students in the primary school take school-level exams in six subjects including Chichewa (the vernacular language) English, mathematics, primary science, social studies, and art and life skills at the end of each semester. Passing the exams in the last semester is required for a student to proceed to the next

[3]SACMEQ stands for Southern African Consortium for Monitoring Educational Quality. For more information, see http://www.sacmeq.org/sites/default/files/sacmeq/reports/sacmeq-iii/policy-brief/mal_achievement_policy_brief_14_ocotberr_2011_latest.pdf

[4]For example, no school among our target schools has electricity in the classroom and only 67% of students have his or her own desk and chair.

grade. Students must pay a fee about USD 0.5 to 1 to take the exam, to cover the print of exam copies. 8th graders take the Primary School Leaving Certificate Exam (PSLCE) to obtain the secondary school admission.

2.2 Program Descriptions

The experiment was implemented in grades 5 to 8 in 31 public primary schools in TA Chimutu,[5] within Lilongwe District in Malawi. TA Chimutu, located about 15km from the capital city of Lilongwe, consists of rural villages and has three school districts. The scholarship programs were implemented by Africa Future Foundation (AFF), an international NGO focused on health and education programs in Malawi and several other countries in Africa.

Project chronology is summarized in Figure 1. Baseline and follow-up survey were implemented during the first and third semester of 2014-2015 academic year, respectively. The baseline, mid-term, and final exam were administered in the end of the first, second, and third semester, respectively, by the NGO and local primary education authorities. The exams are developed by the district level exam committee and therefore exams were same across schools. 8th graders took the PSLCE in the third semester instead of the final exam.

The NGO conducted a baseline exam twice at the end of the first semester and beginning of the second semester. The first and second baseline exam consists of 15 exam questions from six subjects.[6] In order to have more representative sample, those who participated one of two exams (N=8,597, 89.7%) are eligible for the scholarship program.

The exam committee consists of eight teachers, one vice-principal, and one principal (head teacher) of the schools within the district.[7] The exam were jointly administered by the NGO and local primary education authorities. Since scholarship is eligible based on the final exam, we additionally hired teachers, one from each school, as a committee member of the final exam, due to the the fairness concern. The NGO provided exam copies for the students, which exempts students exam fee, during the study period.

[5]TA stands for Traditional Authority and is the administrative division below the level of district.

[6]Only 6728 (70.2%) students among were able to take the first baseline exam due to the exam fee. The NGO support the exam fee in the second baseline exam, and thus 7945 (82.9%) students join the second baseline exam. Mean (and standard deviation) of the first and second exam is very similar: 11.5 (3.2) and 11.5 (3.4), respectively.

[7]Before the NGO's project, each school makes its own exam. The NGO organized an exam committee under the supervision of the local authority to form common questions for the whole region. The school principals recommended mostly experienced teachers in making district level mock Primary School Leaving Certificate Exam (PSLCE) as the members of the committee. , mostly experienced in making district level mock Primary School Leaving Certificate Exam (PSLCE). District school authorities agreed to cover the same materials during the semester to avoid unintended bias due to the school materials.s environment.

The first is scholarship intervention. Specifically, we study two types of merit-based scholarships; *Standard* and *Relative* merit-based scholarship. The design of *Standard* merit-based scholarship program was similar to that in Kremer, Miguel, and Thornton (2009). Within each grade, students scoring in the top 15 percent of all students in the district in the final exam were eligible to receive the award.[8]

The *Relative* merit-based scholarship program was based on the "pay for percentile" design of Barlevy and Neal (2012). In each grade, students were grouped into bins of 100 students by the baseline test score. The top 15 percent of each bin in the final exam were eligible to receive the award. The award consists of a choice among a cash award of 4,500 Malawian Kwacha (about USD 9.70), shoes, a school bag, or school uniform of similar value. The award were distributed in an area-wide awards ceremony that took place in October 2015.[9] A concern on *Relative* merit-based scholarship program is that it could be difficult for 5th to 8th graders to understand the eligibility of the program. Therefore, when the NGO announce randomization results, they had a one-hour session to explain the programs to all three research groups, and administered a quiz to measure students' level of understanding. Figure A1 presents five questions that measures overall understanding of the scholarship programs.

Another student intervention we study is a feedback on student rank. A result of the mid-term exam were randomly provided to students (except for 8th graders) in the beginning of the third semester. Figure 4 shows an example of the feedback note. Figures in the left column present feedback notes given to the feedback treatment group and those in the right column shows feedback notes given to the control group. Figure 4a and Figure 4b compares feedback treatment for the *Standard* merit-based scholarship group. Randomly selected students received their overall rankings in the mid-term exam relative to all students in the program (Figure 4a). Selected students in the *Relative* merit-based scholarship group additionally received information on their rankings in the mid-term relative to students in their respective bins (Figure 4c).

3 Research Design

3.1 Experimental Design

The overall research design is depicted in Figure 2 and Table 1. In February 2015, we stratified the 119 available classrooms within the 31 public primary schools by grade and randomly

[8] For 8th graders, eligibility is determined by PSLCE results.

[9] About 95% of eligible students decided to receive a cash as the award.

assigned classrooms into three groups: the *Standard* merit-based scholarship, the *Relative* merit-based scholarship, or the control group. The results of the scholarship randomization were announced in the middle of the second semester (February 2015).[10] In order to maximize the power of tests between the two scholarship groups, the scholarship groups were over-weighted in the randomization, such that 46 classrooms were assigned to the *Standard* merit scholarship, 43 to the *Relative* merit scholarship, and 30 to the control group.

Figure 3a, 3b, and 3c of Figure 3 shows scholarship program announcement note that were given to students assigned to the *Standard* merit-based scholarship group, the *Relative* merit-based scholarship group, and the control group, respectively. For the *Standard* merit-based scholarship group, information on overall rank as well as scholarship eligibility condition (top 15 %) were provided. For the *Relative* merit-based scholarship group, information on overall rank and rank within bin as well as scholarship eligibility condition (top 15% within bin) were provided. For the control group, only information on overall rank were provided.

Because classrooms were randomly assigned within schools, the program was explained to all children within the classrooms. In each class, program staff explained each program in detail and that the programs were randomly assigned. Students were then told their assigned group, and a short quiz was administered to check the understanding of the program and the students' own program assignments.

After the mid-term exam, students in all three research groups were randomly assigned individually, in equal proportions, to receive feedback treatment. Those selected for the feedback treatment group were received information on their rank in the mid-term exam. 8th graders were excluded from the feedback experiment due to the academic schedule. Figure 4 presents feedback note that students received in the second semester.

What is unique in our setting compared to the previous literature is that we are in an environment where feedback could potentially more effective due to it is linked to the scholarship eligibility. However, on the other hand, students in this study already have information on their previous academic performance which make feedback effect less effective.[11]

[10] the NGO staff had a meeting with school head teachers (principals) to invite schools to actively participate and cooperate with the organization. Head teachers were asked to announce information about the program to the regular school teachers and parents in a school assembly. In the second semester (March and April of 2015), the NGO held additional community meetings to reinforce knowledge about program rules in advance of the final exams.

[11] Overall rank in the baseline was provided in the second semester (through scholarship announcement note in Figure 3) and the third semester (through feedback announcement note in Figure 4).

3.2 Data

We use several sources of data for this research: the district level test score data (the baseline, mid-term, and final exams), student's school attendance data, baseline and follow surveys, and household census.

First, we use is district level test score data. Test score is standardized and student's ranks are determined within each grade. The baseline exam data were used to form bins in the *Relative* merit based scholarship; the mid-term exam data were used for the feedback intervention; and the final exam data were used to measure school achievement and select scholarship recipients. We also collected attendance of students in school through unannounced checks.

Next, the surveys collected information on basic household demographic information, student study habits, parental support for education, student motivation to study, cognitive ability, and non-cognitive traits such as self esteem, grit, and conscientiousness. We collected a list of 9,419 enrolled students in the participating schools during the first semester. Among them, we complete the baseline survey for 7638 students (81%) and 8,491 (90.1%) students participated in the baseline exam. Finally, study sample is 7,386 (78.4%) people who joined in both the baseline survey and baseline exam.

Study hours (per week) is calculated based on self-reported information from the survey. Specifically, we construct study hours measure by multiplying frequency of study per week and average study hours per time. Self esteem is based on Rosenberg self-esteem scale which measures both positive and negative feelings about the self (Rosenberg (1965)). We use short version grit scale to measure grit score (Duckworth and Quinn (2009)). Conscientiousness is one of the Big Five Factors of personality, so it is measured by questions based on Big Five Inventory(BFI) scale (John and Srivastava (1999)).

Teacher effort is measured by students. Students evaluated teachers in eight dimensions - how much teachers cared, challenged, controlled, clarified, conferred, and consolidated student. Parental effort is based on the self-reported information that asks how much parents encourage, help, and ask student to study.

In addition, we measured understanding of the scholarship program and expectation of the scholarship right after the scholarship randomization announcement and at the follow-up survey. To measure the understanding of the scholarship program, students were asked to determine if the scholarship can be awarded in a given situation and to identify the specific condition the student must satisfy to win scholarship (Figure A1). We also asked perceived probability of receiving the scholarship.

Our study sample is 7,385 students from 31 primary schools who participated in both baseline survey and exam (the 2014 end-of-year exams). Column (1) of Table 2 displays summary statistics of key variables. Average age is 14.2 and about 47.3 percent of the sample are males. The asset index is created by using principal component analysis and standardized. The test score is standardized. As for the baseline survey, attendance rate of the students is 85% and the average study hours per week is 16.1.

Columns (3) and (4) test of differences in means across scholarship treatment groups. In addition, column (6) present the differences in feedback treatment and control group. Overall, we observe few significant differences. Of the 15 variables examined, only one variable between *Standard* merit and control group is significantly different. In feedback randomization four out of 15 are significantly different in 10% level, but the differences are not economically significant. For example, the difference in Grit is only 0.64% (=0.02/3.18) compared to overall average score. When we compare the scholarship programs with the control group, one of the 20 (5%) individual comparisons are significant at the 10 percent level. We also see few differences between the feedback and no-feedback groups: only one variable out of 16 is significantly different at the 10 percent level.

Table A3 displays sample attrition across treatment groups. On average 88, 83, and 90 % of study sample joined the mid-term exam, follow-up survey, and final exam, respectively. We observe no significant differences between scholarship groups and the control group, and no significant difference between the feedback treatment and the control group. In general, we do not find evidence of systematic attrition, but Column (3) shows that those who are top 15% in the final exam among the scholarship treatment groups are less likely to participated in the follow-up survey than top 15% in the control group. Therefore the results must be interpreted with this caveat.

4 Results

4.1 Estimating Equation

We employ a number of empirical strategies to estimate impacts of being assigned to scholarship programs. First, we estimate the following equation:

$$Y_{ijgk} = \beta_0 + \beta_1 Standard_{ij} + \beta_2 Relative_{ij} + \eta_g + \gamma_k + X_{ijgk} + \varepsilon_{ijgk} \quad (1)$$

where Y_{ijlk} is the outcome such as rank and score of the final test, expectation of the scholarship, student motivation to study, student study habits, student non-cognitive traits such as self esteem,

grit, and conscientiousness, and parental support for education for student i in classroom j of grade g and district k. *Standard* and *Relative* is an indicator of being *Standard* and *Relative* merit-based scholarship group, respectively. η is a grade fixed effect and γ is district fixed effect. The control vector, X, includes age, race, household size, and asset index. Standard errors are clustered at the the classroom level.

Because the distributional impacts of the programs is a key research question, we also interact the treatment groups with an indicator for whether the student's baseline rank was in the top 15 percent:

$$\begin{aligned} Y_{ijgk} =& \beta_0 + \beta_1 Standard_{ij} + \beta_2 Relative_{ij} + \beta_3 Top15_{ijgk} \\ &+ \beta_4 Standard_{ij} * Top15_{ijgk} + \beta_5 Relative_{ij} * Top15_{ijgk} + \eta_g + \gamma_k + X_{ij} + \varepsilon_{ijk} \end{aligned} \tag{2}$$

Where *Top15* is an indicator of being within top 15 percent in the baseline test. In these specifications, β_4 and β_5 captures whether top 15 percent students assigned to the *Standard* and *Relative* merit-based scholarship group respond differently compared to top 15 percent students in the control group.

Next, to analyze the impacts of feedback, we utilize the following equation:

$$\begin{aligned} Y_{ijgk} =& \beta_0 + \beta_1 Standard_{ij} + \beta_2 Relative_{ij} + \beta_3 Feedback_{ijg} + \beta_4 Standard_{ij} * Feedback_{ijg} \\ &+ \beta_5 Relative_{ij} * Feedback_{ijg} + \eta_g + \gamma_k + X_{ijgk} + \varepsilon_{ijgk} \end{aligned} \tag{3}$$

where *Feedback* indicates student i 's assignment to receive feedback.

4.2 Understanding of Program and Expectation of Scholarship

Before turning to the main impact results, we first discuss students' understanding of the program and expectation that they would receive the scholarship. As described above, the program announcement and follow-up surveys included a quiz that tested the understanding of the scholarship program for students in all treatment and control groups. The quiz contained 5 questions about students who were hypothetically assigned to one of the scholarship groups and whether they would receive the scholarship given their absolute or relative rank in the program.

As shown in Columns 1 and 2 of Table 3, students understood the scholarship program quite well.

For example, students answered 92 percent of questions correctly at the time of the program announcement, and the average falls to 64 percent as of the follow-up survey. Panel A shows that there are no significant differences between treatment groups at either baseline or follow up, with confidence intervals able to rule out differences above about five percentage points. Examining differences in understanding by baseline exam performance in Panel B, there is no significant difference in understanding between top 15 percent students and lower 85 percent students right after the announcement (Column 1) , but top 15 percent students understand better amount scholarship mechanism than lower 85 percent students by eight percentage points in the follow-up survey (Column 2).

Next, Figure 5 and Columns 3 and 4 of Table 3 display the students' expectation of receiving the scholarship across research groups. If students understand scholarship eligibility well, students in the *Standard* merit-based scholarship group with low/high baseline test score should have low/high expectation; for students in the *Relative* merit-based scholarship group, expectation should not be related with baseline test score; and for students in the control group, expectation should be zero. Figure 5 confirms this pattern, especially when the program announced. Formal regression results in Columns 3 and 4 show that students in the scholarship groups were 29-35 percentage points more likely to expect the scholarship. Examining differences across baseline test scores, those in the top 15 percent in the *Standard* merit-based scholarship group were significantly more likely to expect the scholarship in both rounds, with differences of 43 and 13 percentage points, respectively. In sum, analysis on understanding and expectation confirms that the treatment went well as we expected.

4.3 Impacts on Test Score

Figure 7 non-parametrically illustrates the impact of two scholarship programs on the final exam rank and score. The baseline and final test rank are displayed in X and Y axis, respectively. The endline ranks and scores for the *Standard* merit-based group are lower than those of the control group for all baseline scores except those with the highest baseline scores, above about the 90th percentile. On the other hand, impacts of the *Relative* merit-based scholarship are positive for students with lower baseline scores and negative for higher baseline scores.

Table 4 presents OLS estimates of the impacts of the scholarship schemes on students rank in the final exam (Columns 1 and 2), standardized overall score in the final exam (Columns 3 and 4), standardized math score in the final exam (Columns 5 and 6), and standardized math score measured in the final survey (Columns 7 and 8).[12] In addition, Panel A and B present results from

[12]In each outcome, we present two specifications, but the results are robust in various specifications.

equations (1) and (2), respectively. As shown in Columns 1 and 3, the impact of the *Standard* merit-based scholarship is significantly *negative*: on average, students in this group were seven percent lower in rank and 0.26 standard deviations lower in the test score than the control group. Direction of the impacts are same in the math tests of the final exam and test in the survey as shown in Columns 5 and 7. Panel B confirms the pattern that the decrease in academic achievement are driven by students with initial test scores in the bottom 85 percent: the coefficient on *Standard* merit scholarship is negative and significant, and that on the interaction between *Standard* merit scholarship and baseline top 15 percent in equation (2) is of opposite sign and almost equal in magnitude to main effect of the scholarship.

By contrast, the impacts of the *Relative* merit-based scholarship are smaller and not statistically significant. On average, students in this group were four percentage points lower in rank at endline than the control group and 0.12 standard deviations lower, and neither estimate is statistically significant. Turning to interactions with initial test score, it appears that the top students in this group scored lower than the top students in the control group (by 4.3 percentage points in rank and 0.32 standard deviations), but again neither estimate is statistically significant.

4.4 Intermediate Outcomes I: Student's Response

In this subsection we analyze intermediate outcomes such as students efforts, motivation, and non-cognitive traits that could have served as mechanisms for the test rank score results presented in Section 4.3. Table 5 analyzes students' effort including attendance and study hours (Columns 1 and 2), motivation (Column 3), and non-cognitive traits including self-esteem, grit, and conscientiousness (Columns 4-6). Figure 8 and Figure 9 are corresponding results non-parametrically.

As shown in Column 1, there are no significant differences in attendance between the *Standard* or *Relative* merit-based scholarship group and the control group. In addition, is no evidence for heterogeneity by baseline test score as shown in Panel B. We find similar null impacts on self-reported weekly study hours measured in the follow-up survey as shown in Column 2, but point estimates suggest slightly less study effort in both scholarship treatment groups (Panel A), and slightly lower effort among students with the highest baseline scores (Panel B). However, neither the average effects or effects by baseline test score are statistically significant.

Columns 3 to 6 present impacts on non-cognitive measures including motivation to study, self esteem, grit, and conscientiousness. In general, scholarship program negatively affect student non-cognitive measures as shown in Panel A. Negative impacts on non-cognitive measures are generally larger in the *Standard* merit-based scholarship program and mirror the test score results. Across

the measures of motivation, self esteem, and conscientiousness, we find statistically significant and negative impacts for the *Standard* merit scholarship program relative to control group. In addition, these impacts are concentrated among the bottom 85 percent of students as shown in Panel B and Figure 9. By contrast, the negative impacts of of the *Relative* merit-based scholarship program are much smaller and statistically insignificant, with an exception of self-esteem, over the full sample or by baseline test score.

4.5 Intermediate Outcomes II: Teacher and Parental Response

We also explore teacher and parental responses to the scholarship programs. Table 6 presents impacts on students' perceptions of teacher and parental effort. We do not find evidence on change in student's perception on teacher and parental efforts. However, we find that parents mentioned the scholarship program more often especially in the *Standard* merit-based scholarship group, with effects concentrated among children with the highest baseline test scores. By contrast, in the *Relative* merit-based scholarship group, parents response is smaller and insignificant, and not related with baseline test score.

4.6 Feedback

Lastly, we study how feedback on rank affect student performance. Figure 11 presents the final exam rank by mid-term exam rank for whole sample (Figure 11a) as well as by scholarship group (Figure 11b). Figure 11ashows that for those ranked top 15 percent in the feedback treatment group performed better than those in the control group. Table 7 presents formal regression results. We estimate equation (3) for the full sample (Columns 1 and 4), as well as for the top 15 percent (Columns 2 and 5) and bottom 85 percent of students at baseline (Columns 3 and 6). Although we do not find significant difference between the feedback treatment and control group, Panel A of Column 2 confirms the pattern shown in Figure 11 that feedback is effective only for the top performers.

5 Conclusion

We study the impacts of two types of merit-based scholarships as well as feedback of academic performance for 5th to 8th graders in Malawi. One criticism of merit-based scholarship programs is that by providing rewards to only the very top performers, lower-performing students who are

unlikely to receive the incentive may not respond to the programs. An incentive design that could address this concern follows that proposed by Barlevy and Neal (2012), in which students are grouped by baseline score, and incentives are awarded to the top performers in each group. We study the impacts of this *Relative* merit-based scholarship program alongside a more typical *Standard* merit-based scholarship program.

We implement a field experiment where 119 classrooms in 31 schools were randomly assigned to one of the three groups: *Relative* merit-based scholarship program, *Standard* merit-based scholarship program, and the control group. For those in the *Standard* merit-based scholarship program, top performers in the district in the final exam are awarded. An incentive design of *Relative* merit-based scholarship program in which students are grouped by baseline test score, and incentives are awarded to the top performers in each group. Another student intervention we study is a feedback on student rank. A result of the mid-term exam were randomly provided to students in the middle of the study period.

We find that the *Standard* merit-based scholarship significantly decreased test scores compared to the control group, with the largest decreases concentrated among those least likely to win the scholarship. These decreases in test scores correspond to decreases in motivation to study among those least likely to win. In addition, we find feedback on ranking may improve students performance only for high performers where award is given to the high performers.

References

Azmat, Ghazala and Nagore Iriberri (2010). "The importance of relative performance feedback information: Evidence from a natural experiment using high school students". *Journal of Public Economics* 94.7-8, pp. 435–452.

Bandiera, Oriana, Valentino Larcinese, and Imran Rasul (2015). "Blissful ignorance? A natural experiment on the effect of feedback on students' performance". *Labour Economics*. European Association of Labour Economists 26th Annual Conference 34, pp. 13–25.

Barlevy, Gadi and Derek Neal (2012). "Pay for Percentile". *American Economic Review* 102.5, pp. 1805–1831.

Behrman, Jere R., Susan W. Parker, Petra E. Todd, and Kenneth I. Wolpin (2015). "Aligning Learning Incentives of Students and Teachers: Results from a Social Experiment in Mexican High Schools". *Journal of Political Economy* 123.2, pp. 325–364.

Berry, James (2015). "Child Control in Education Decisions An Evaluation of Targeted Incentives to Learn in India". *Journal of Human Resources* 50.4, pp. 1051–1080.

Blimpo, Moussa P. (2014). "Team incentives for education in developing countries: A randomized field experiment in Benin". *American Economic Journal: Applied Economics* 6.4, pp. 90–109.

Dizon-Ross, Rebecca (2016). "Parents' Beliefs and Children's Education: Experimental Evidence from Malawi". *Mimeo, University of Chicago*.

Duckworth, Angela Lee and Patrick D. Quinn (2009). "Development and validation of the short grit scale (grit-s)". *Journal of Personality Assessment* 91.2, pp. 166–174.

Gneezy, Uri, Stephan Meier, and Pedro Rey-Biel (2011). "When and Why Incentives (Don't) Work to Modify Behavior". *Journal of Economic Perspectives* 25.4, pp. 191–210.

Hirshleifer, Sarojini (2015). "Incentives for Effort or Outputs? A Field Experiment to Improve Student Performance". Tech. rep. Working Paper.

John, Oliver P. and Sanjay Srivastava (1999). "The Big Five trait taxonomy: History, measurement, and theoretical perspectives". *Handbook of personality: Theory and research* 2.1999, pp. 102–138.

Kremer, Michael, Edward Miguel, and Rebecca Thornton (2009). "Incentives to Learn". *Review of Economics and Statistics* 91.3, pp. 437–456.

Lazear, Edward P. and Sherwin Rosen (1981). "Rank-Order Tournaments as Optimum Labor Contracts". *Journal of Political Economy* 89.5, pp. 841–864.

Li, Tao, Li Han, Linxiu Zhang, and Scott Rozelle (2014). "Encouraging classroom peer interactions: Evidence from Chinese migrant schools". *Journal of Public Economics* 111, pp. 29–45.

Loyalka, Prashant Kumar, Sean Sylvia, Chengfang Liu, James Chu, and Yaojiang Shi (2016). "Pay by Design: Teacher Performance Pay Design and the Distribution of Student Achievement".

Rosenberg, Morris (1965). "Society and the Adolescent Self-Image." *Science* 148.3671, pp. 804–804.

Sharma, Dhiraj (2010). "The impact of financial incentives on academic achievement and household behavior: Evidence from a randomized trial in Nepal".

Tran, Anh and Richard Zeckhauser (2012). "Rank as an inherent incentive: Evidence from a field experiment". *Journal of Public Economics* 96.9-10, pp. 645–650.

Figure 1: Project Chronology

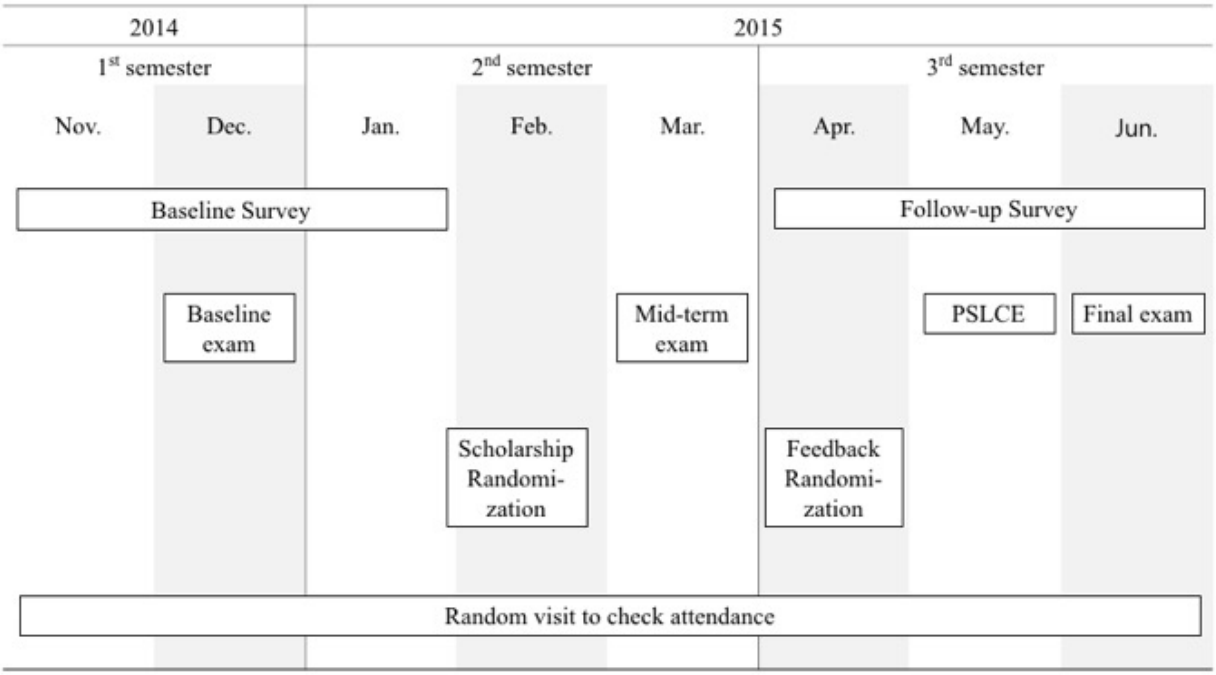


Figure 2: Experimental Design

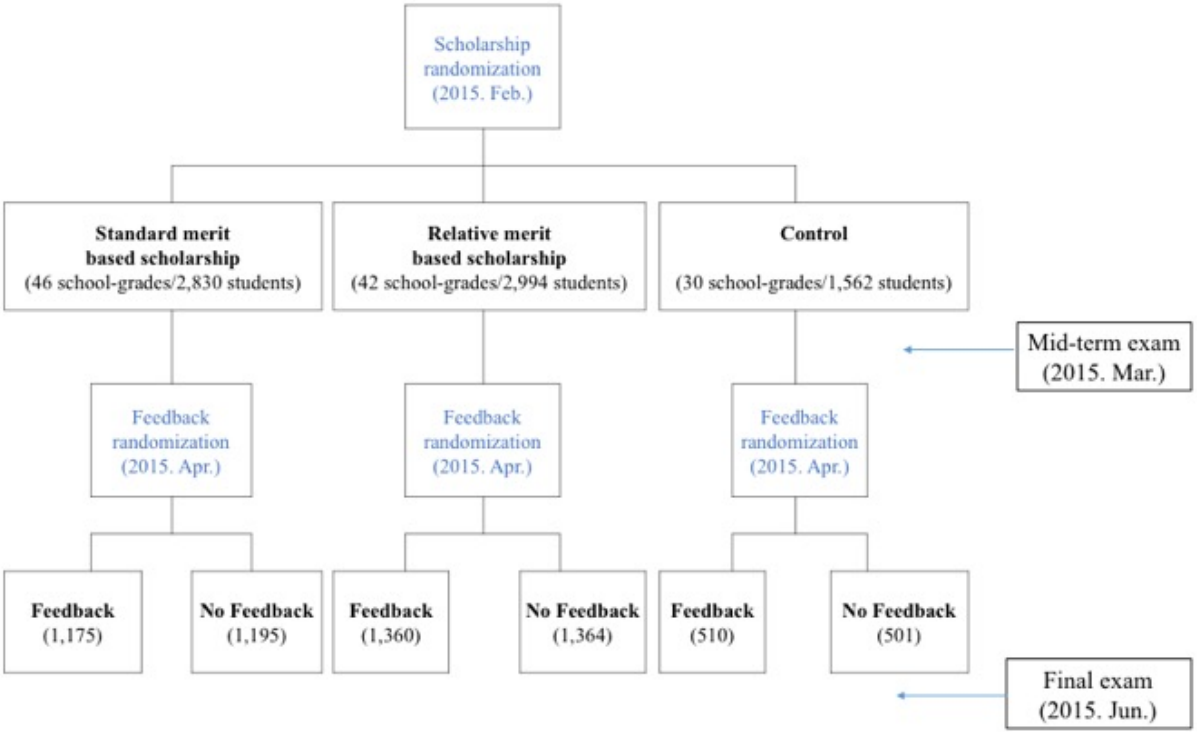
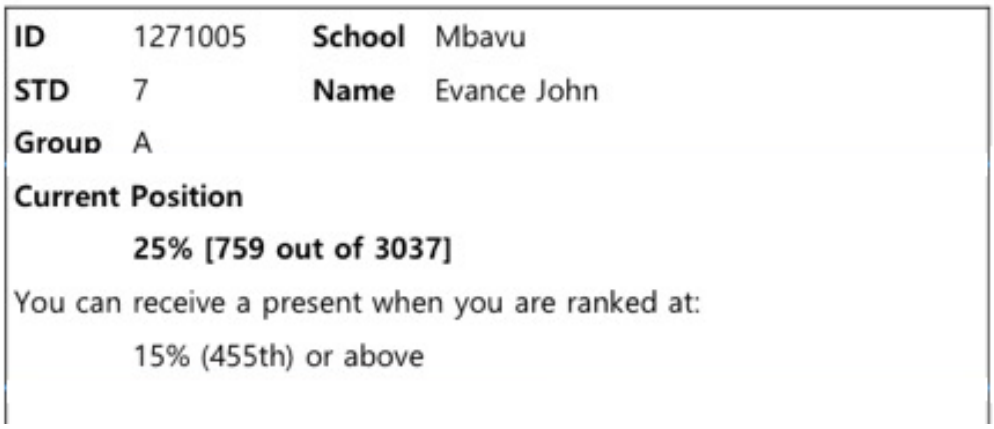
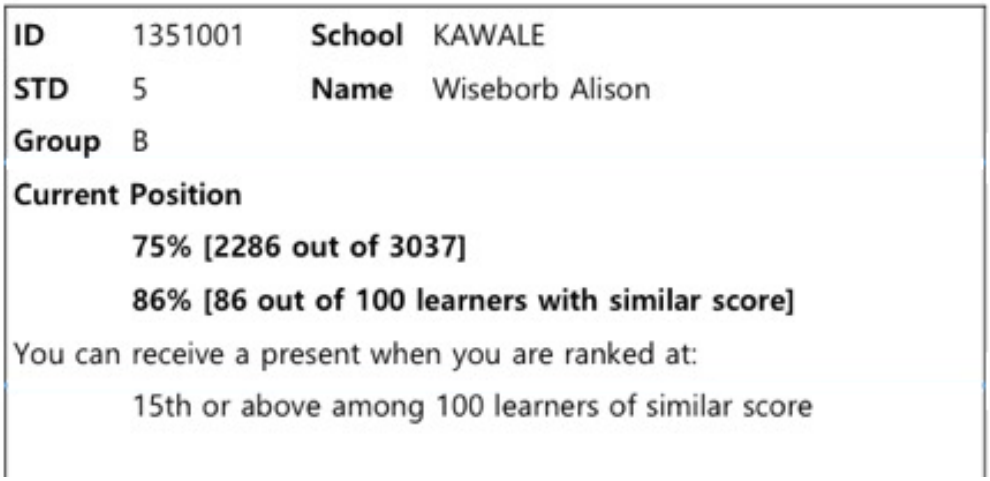


Figure 3: Scholarship Randomization result announcement note

(a) *Standard* merit-based scholarship group



(b) *Relative* merit-based scholarship group



(c) Control group

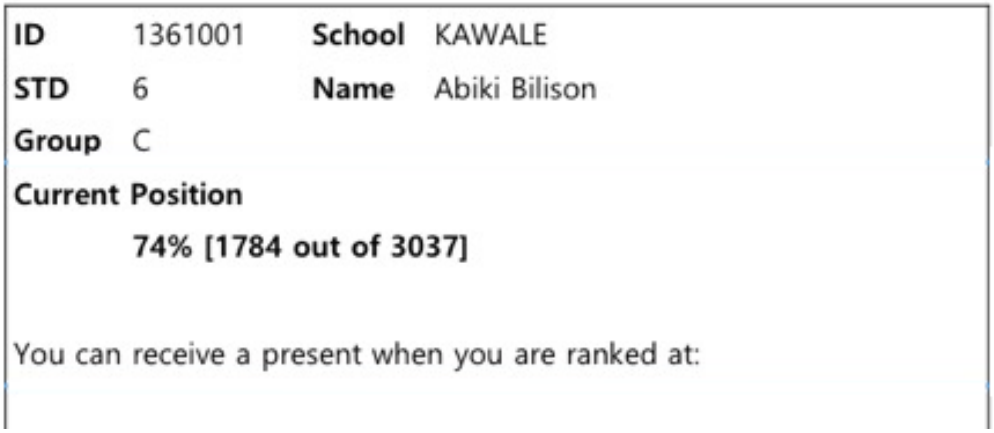
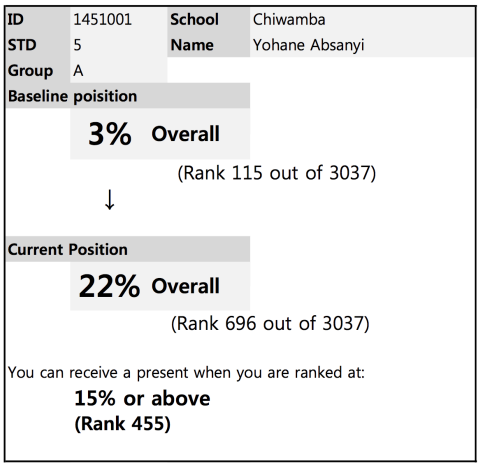
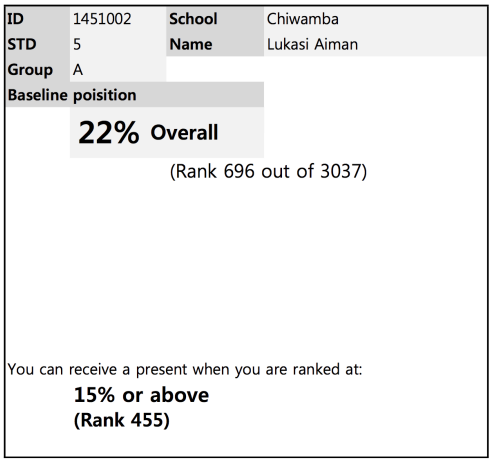


Figure 4: Feedback note

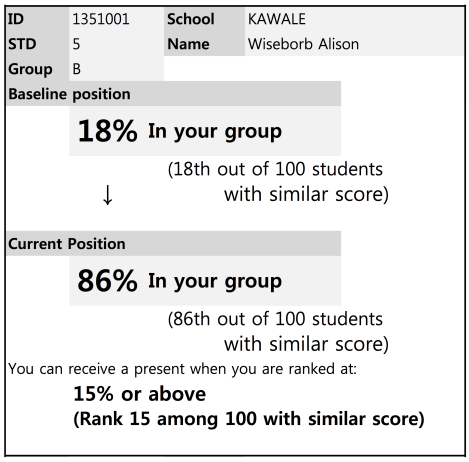
(a) Feedback and *Standard*



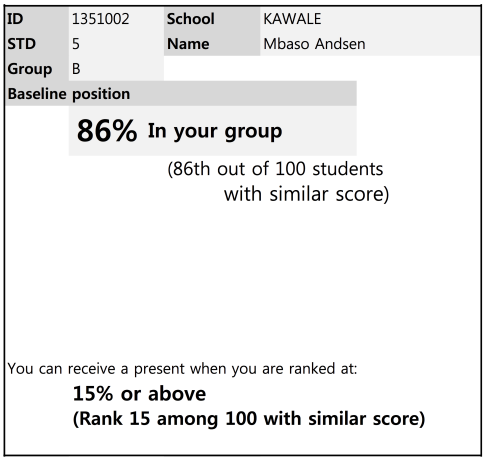
(b) No Feedback and *Standard*



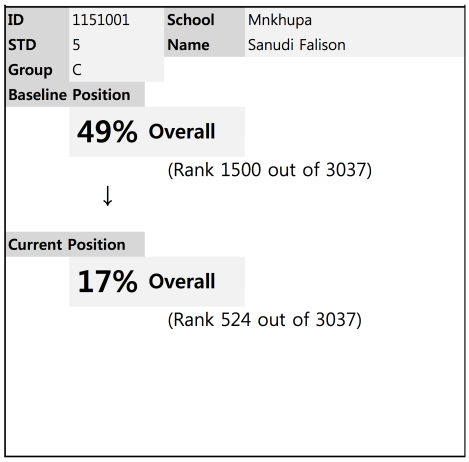
(c) Feedback and *Relative*



(d) No Feedback and *Relative*



(e) Feedback and Control



(f) No Feedback and Control

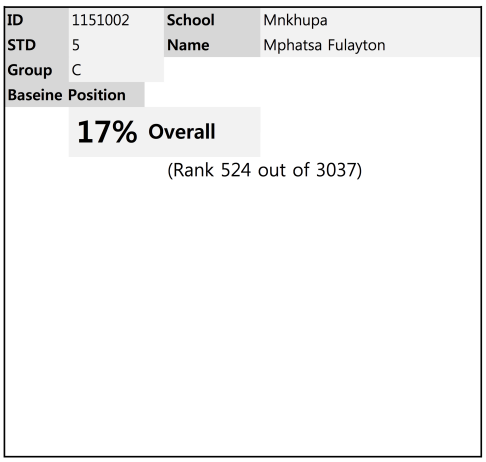


Figure 5: Understanding of the program

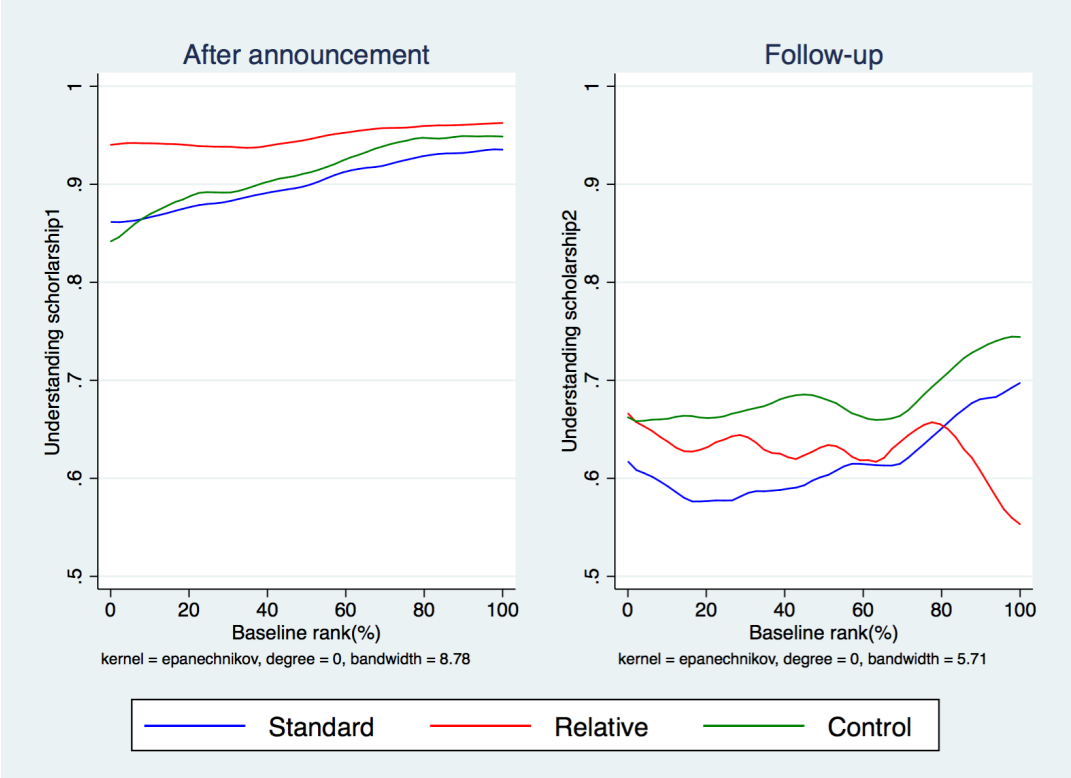


Figure 6: Expectation of Scholarship

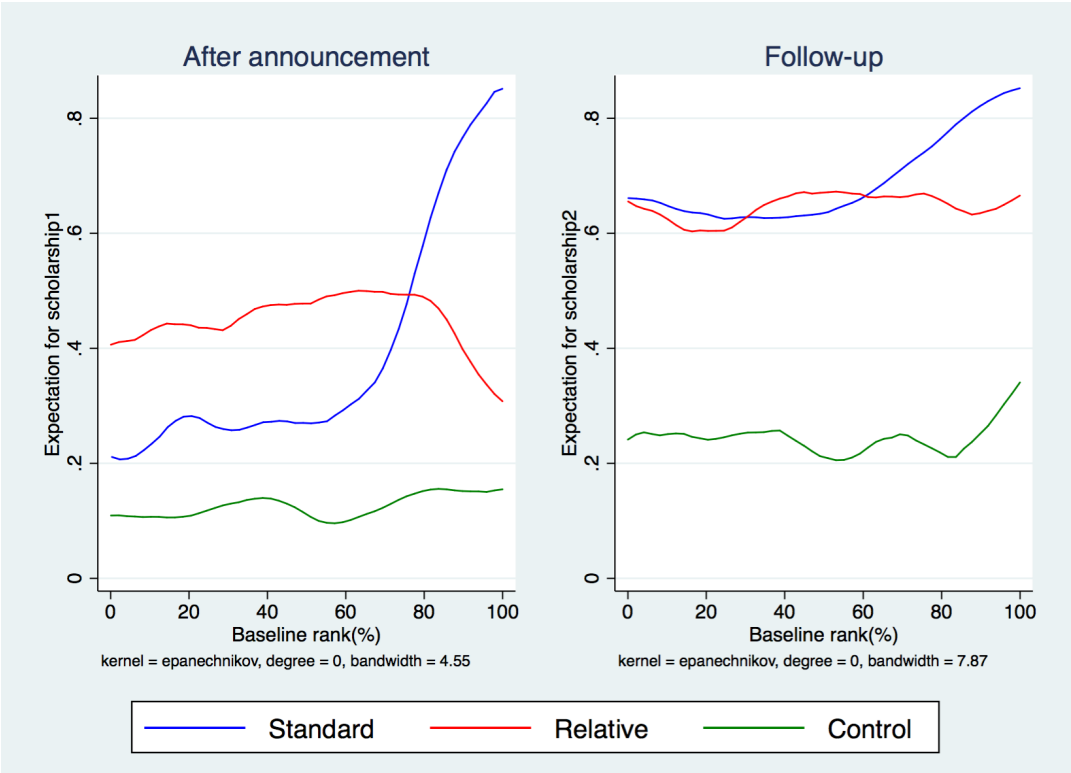


Figure 7: Exam results, by Baseline Rank

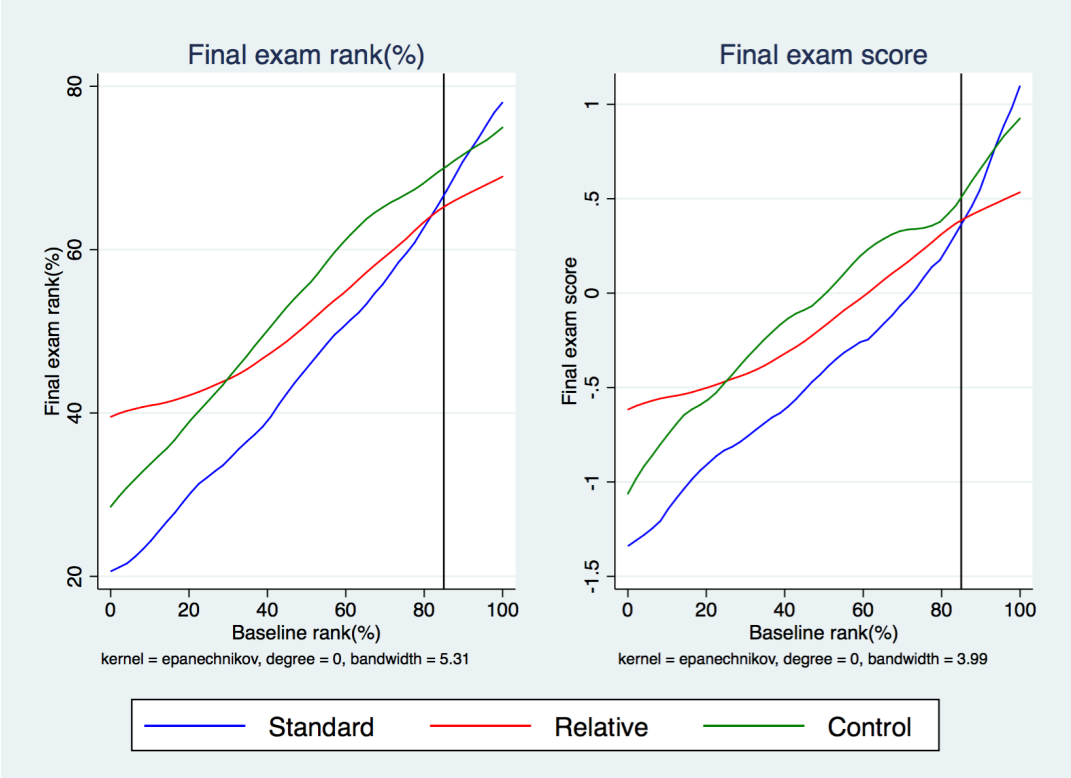


Figure 8: Student input, by Baseline Rank

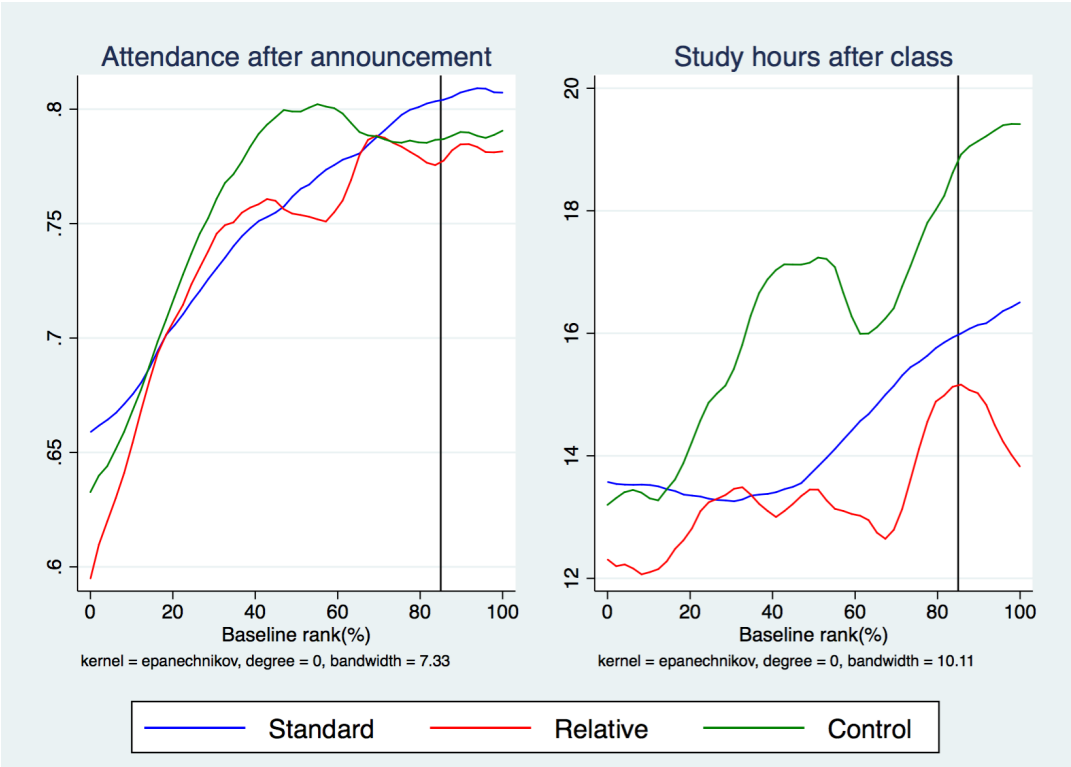


Figure 9: Motivation and self-esteem, by Baseline Rank

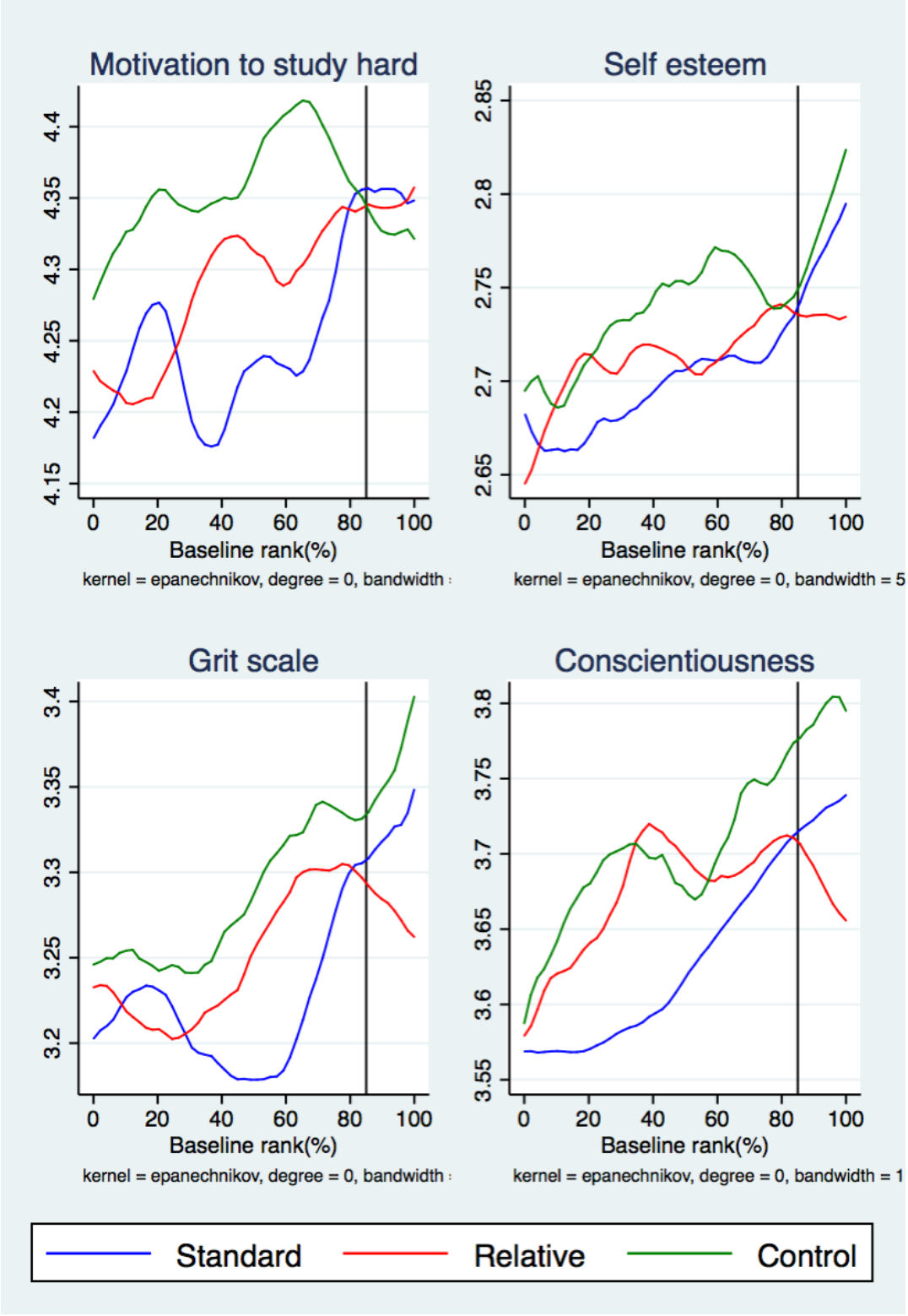


Figure 10: Teachers and Parental response, by Baseline Rank

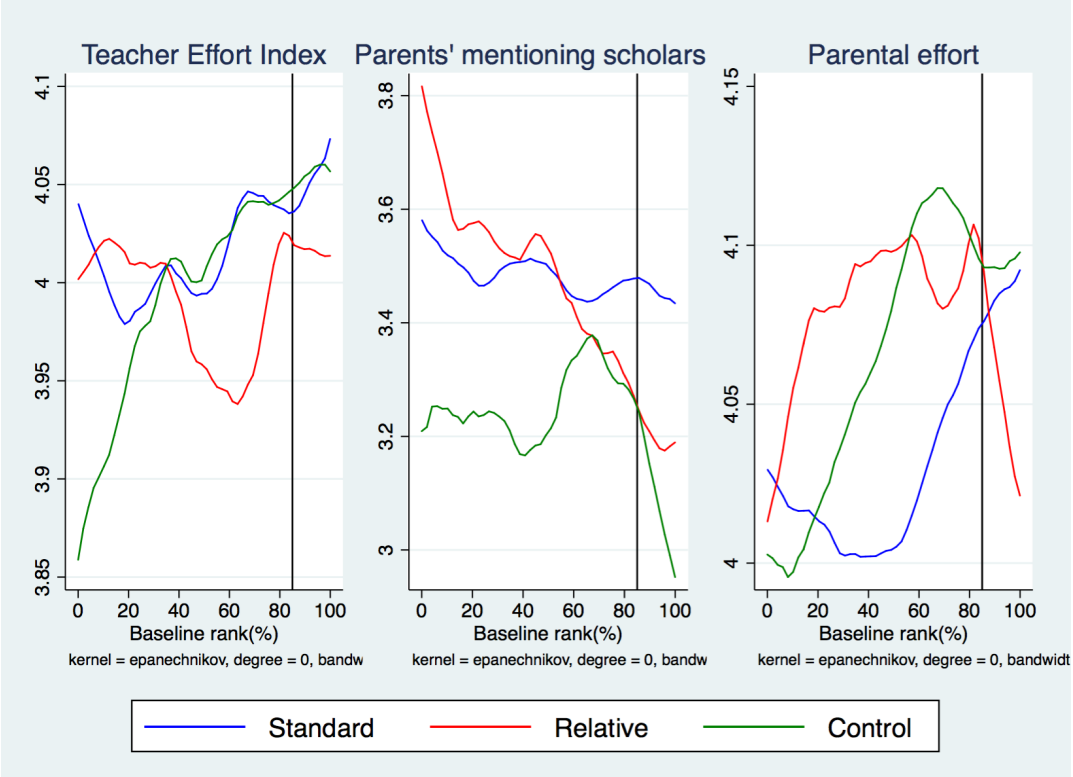
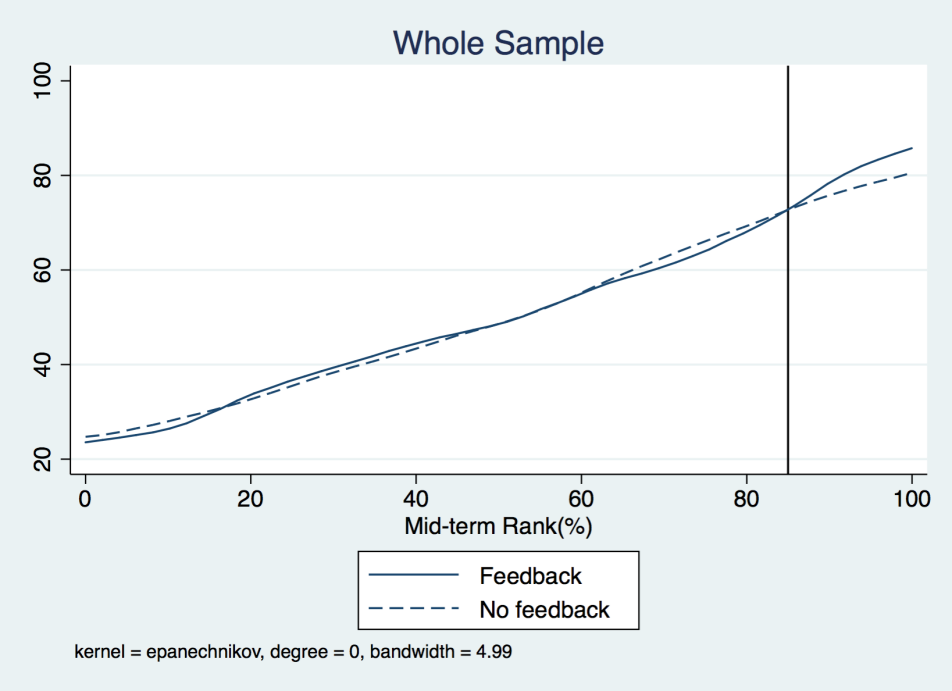


Figure 11: Exam results and feedback

(a) Whole sample



(b) By treatment group

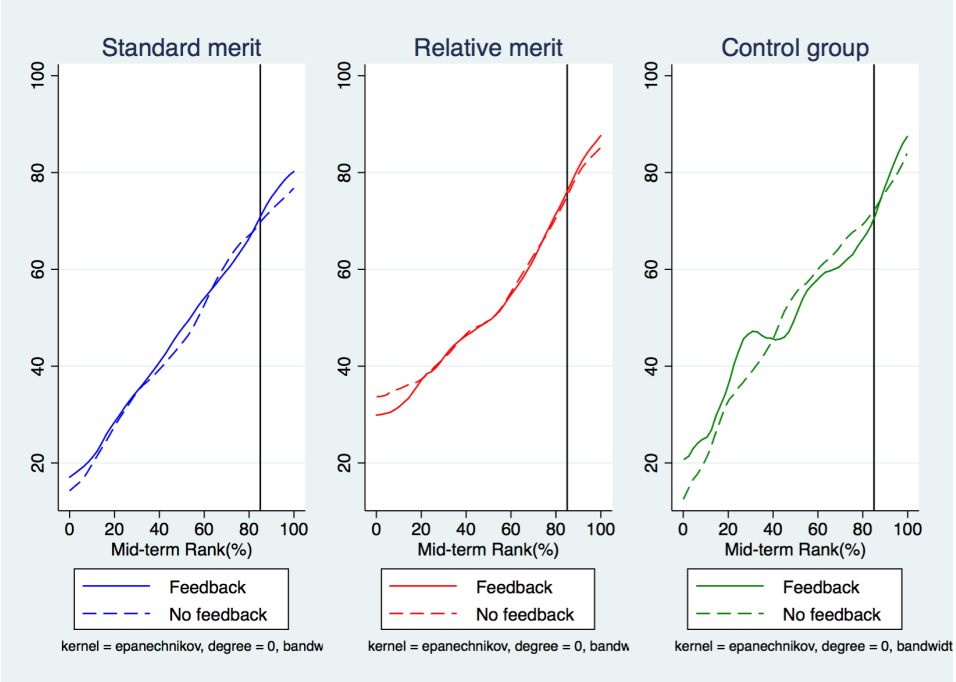


Figure 12: Exam results, by Baseline Rank

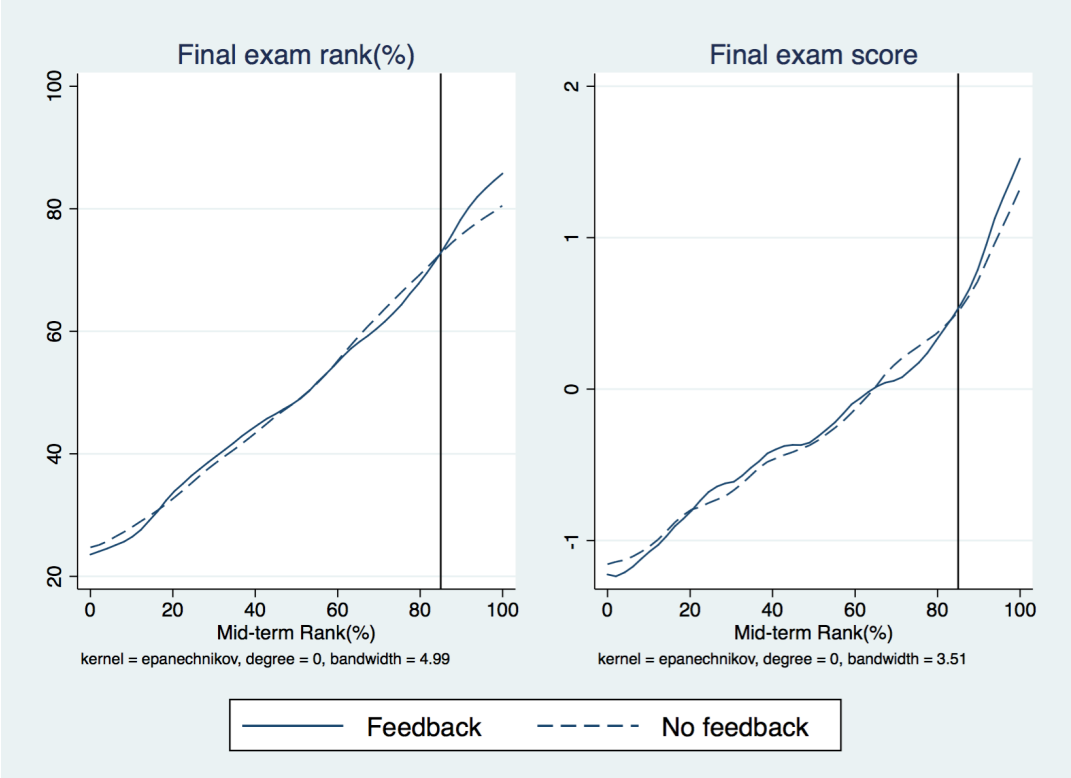


Figure 13: Student input, by Baseline Rank

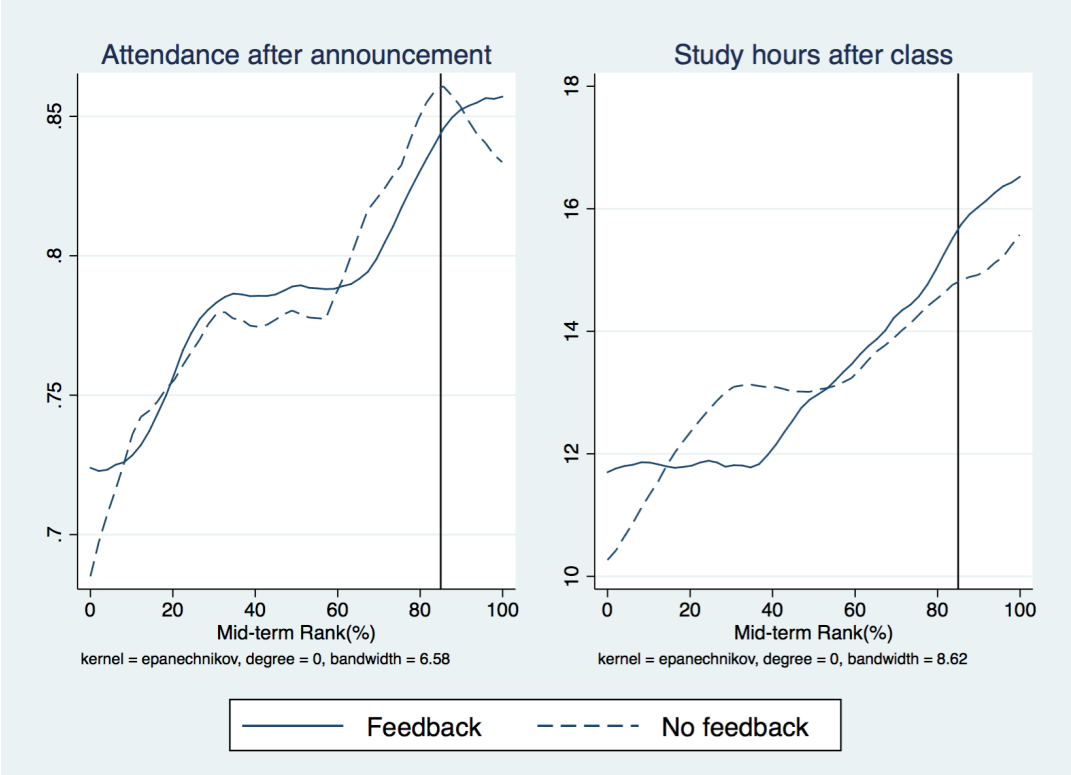


Figure 14: Motivation and self-esteem, by Baseline Rank

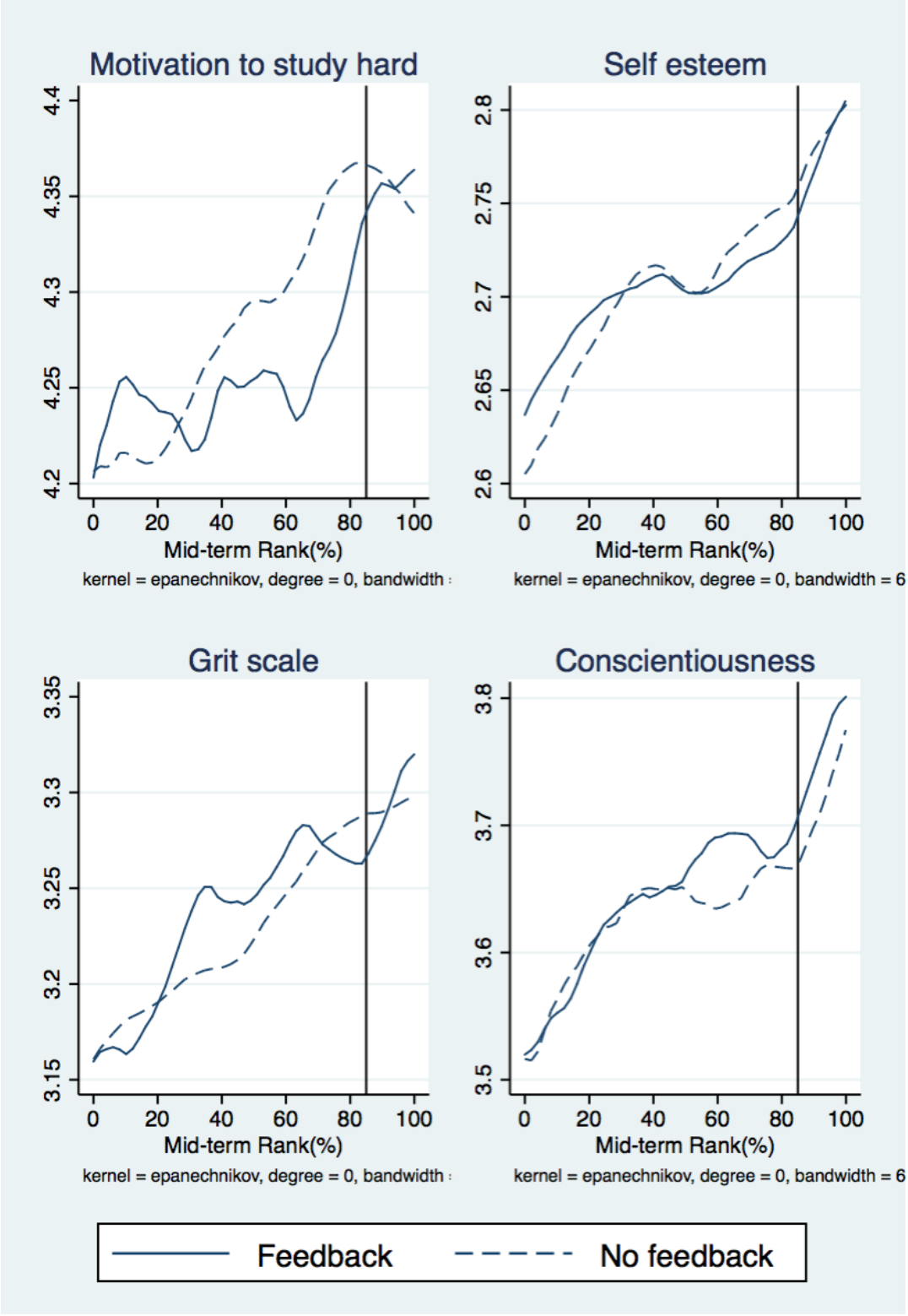


Figure 15: Teachers and Parental response, by Baseline Rank

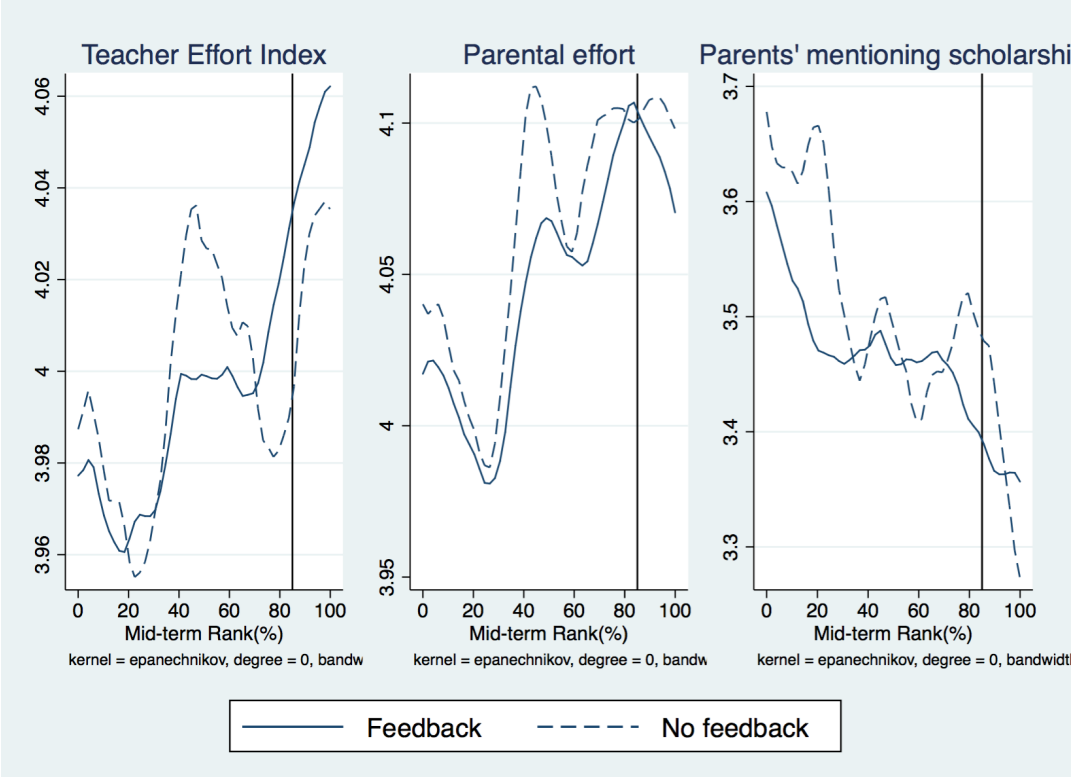


Table 1: Sample Composition by Treatment Category

Scholarship Assignment	Group	Classrooms	Students	Feedback Assignment	Group	Students
Standard merit-based	G1	46	2,830	No Feedback	M0	1,175
				Feedback	M1	1,195
Relative merit-based	G2	43	2,994	No Feedback	R0	1,360
				Feedback	R1	1,364
Control	G3	30	1,562	No Feedback	C0	510
				Feedback	C1	501
Total		119	7,386			6,105

Table 2: Balance of Baseline Variables Across Treatment Groups

		Scholarship Randomization				Feedback Randomization	
	Whole Sample Mean	Control Mean	*Standard* vs. Control	*Relative* vs. Control	N	Feedback vs. No Feedback	N
	(1)	(2)	(3)	(4)	(5)	(6)	(7)
Age	14.2	14.4	-0.366	-0.300	7385	0.199**	6103
	[4.60]	[3.60]	(0.311)	(0.280)		(0.0932)	
Male	0.473	0.486	-0.00358	-0.0275	7385	0.0128	6103
	[0.499]	[0.500]	(0.0195)	(0.0177)		(0.0129)	
Ethnic group: Chewa	0.887	0.914	-0.0329	-0.0360	7358	-0.00274	6077
	[0.317]	[0.280]	(0.0352)	(0.0352)		(0.00641)	
Household size	7.84	7.70	0.164	0.192	7497	0.0782*	6199
	[1.82]	[1.88]	(0.345)	(0.318)		(0.0411)	
Asset index	-0.00396	-0.00919	0.000625	0.0124	7102	-0.0902*	5848
	[1.92]	[1.88]	(0.183)	(0.175)		(0.0510)	
Baseline rank(%)	51.2	50.8	-0.0178	1.15	7497	-0.253	6199
	[28.4]	[27.9]	(3.01)	(4.01)		(0.625)	
Baseline score: Total	-0.00998	0.00000	-0.0543	0.0273	7497	-0.00715	6199
	[1.04]	[0.999]	(0.107)	(0.160)		(0.0204)	
Baseline score: Math	0.0250	0.0282	-0.00528	-0.00285	7407	-0.0105	6112
	[0.982]	[0.961]	(0.0797)	(0.0956)		(0.0230)	
Attendance	0.846	0.858	-0.0127	-0.0177	7497	0.00565	6199
	[0.197]	[0.201]	(0.0181)	(0.0179)		(0.00492)	
Study hours per week	16.1	16.8	-1.00	-0.818	7308	0.163	6031
	[16.1]	[16.4]	(0.865)	(0.871)		(0.374)	
Motivation to study [1-5]	4.52	4.53	-0.0541	0.0159	7374	-0.000297	6092
	[0.811]	[0.789]	(0.0650)	(0.0547)		(0.0210)	
Self-esteem [1-4]	2.65	2.67	-0.0273	-0.0188	7368	0.0105	6087
	[0.336]	[0.338]	(0.0232)	(0.0236)		(0.00688)	
Grit [1-5]	3.18	3.21	-0.0496*	-0.0287	7368	0.0205*	6087
	[0.433]	[0.450]	(0.0256)	(0.0280)		(0.0116)	
Conscientious [1-5]	3.59	3.58	-0.0279	0.0454	7370	0.00236	6089
	[0.586]	[0.600]	(0.0676)	(0.0663)		(0.0154)	
Teacher effort index [1-5]	4.03	3.96	0.0661	0.115	7364	0.00157	6083
	[0.537]	[0.567]	(0.0816)	(0.0724)		(0.0132)	
Parental encouragement	4.44	4.47	-0.0528	-0.0362	7281	0.0393**	6024
	[0.801]	[0.754]	(0.0566)	(0.0483)		(0.0187)	

* denotes significance at 0.10; ** at 0.05; *** at 0.01.

Table 3: Understanding and Expectation

	Understanding of Scholarship		Expectation of Scholarship	
	After Announcement	Follow-up	After Announcement	Follow-up
	(1)	(2)	(3)	(4)
Panel A				
Standard	-0.00778	-0.0194	0.305***	0.442***
	(0.0215)	(0.0220)	(0.0556)	(0.0439)
Relative	0.0346*	-0.0254	0.354***	0.397***
	(0.0200)	(0.0242)	(0.0663)	(0.0455)
R-Squared	0.067	0.098	0.115	0.146
Panel B				
Standard	-0.00826	-0.0139	0.237***	0.417***
	(0.0239)	(0.0223)	(0.0576)	(0.0475)
Relative	0.0384*	-0.00509	0.381***	0.404***
	(0.0221)	(0.0250)	(0.0654)	(0.0486)
Baseline top 15%	0.0157	0.0838***	-0.0647	-0.0429
	(0.0175)	(0.0251)	(0.0468)	(0.0391)
Standard x Top 15%	0.00130	-0.0373	0.452***	0.146***
	(0.0239)	(0.0361)	(0.0775)	(0.0493)
Relative x Top 15%	-0.0230	-0.120***	-0.132	-0.0418
	(0.0215)	(0.0320)	(0.0802)	(0.0520)
Grade fixed	Yes	Yes	Yes	Yes
District fixed	Yes	Yes	Yes	Yes
Demographic control	Yes	Yes	Yes	Yes
Further control	Yes	Yes	Yes	Yes
N	5617	5851	5594	5750
R-Squared	0.068	0.102	0.155	0.150
Mean of Dep. Var.	0.924	0.636	0.356	0.579

Notes: Standard errors are clustered at the the classroom level. * denotes significance at 0.10; ** at 0.05; *** at 0.01.

Table 4: Test Score Impacts

	Final Exam Rank(pct)		Final Exam Score(Norm)		Final Math score(Norm)		Math Score(svy)	
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Panel A								
Standard	-7.876*	-7.607*	-0.280*	-0.271*	-0.222	-0.222	-0.0250	-0.0213
	(4.041)	(3.932)	(0.153)	(0.149)	(0.142)	(0.141)	(0.0194)	(0.0187)
Relative	-4.290	-4.348	-0.127	-0.123	0.0710	0.0713	-0.00732	-0.00639
	(4.573)	(4.464)	(0.182)	(0.178)	(0.160)	(0.159)	(0.0212)	(0.0203)
R-Squared	0.284	0.302	0.291	0.313	0.103	0.115	0.309	0.321
Panel B								
Standard	-8.712**	-8.536**	-0.307*	-0.302*	-0.209	-0.208	-0.0234	-0.0202
	(4.326)	(4.172)	(0.160)	(0.154)	(0.155)	(0.155)	(0.0198)	(0.0187)
Relative	-3.911	-4.001	-0.0789	-0.0789	0.0745	0.0787	-0.0129	-0.0119
	(5.022)	(4.871)	(0.195)	(0.189)	(0.175)	(0.174)	(0.0223)	(0.0211)
Standard x Top 15%	5.146	5.495	0.162	0.178	-0.113	-0.118	-0.00538	-0.00307
	(5.531)	(5.297)	(0.231)	(0.223)	(0.178)	(0.173)	(0.0243)	(0.0249)
Relative x Top 15%	-2.640	-2.450	-0.279	-0.256	-0.126	-0.139	-0.00125	0.000998
	(6.083)	(5.936)	(0.269)	(0.260)	(0.256)	(0.259)	(0.0251)	(0.0252)
Baseline top 15%	2.873	2.889	0.0428	0.0566	0.402***	0.390***	0.147***	0.142***
	(4.329)	(4.183)	(0.189)	(0.185)	(0.143)	(0.139)	(0.0161)	(0.0162)
Grade fixed	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
District fixed	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Baseline value	Yes	Yes	Yes	Yes	Yes	Yes	No	No
Demographic control	No	Yes	No	Yes	No	Yes	No	Yes
N	6689	6353	6689	6353	6585	6253	6171	5857
R-Squared	0.287	0.306	0.295	0.317	0.113	0.124	0.353	0.363
Mean of Dep. Var.	51.258	51.550	-0.156	-0.142	0.023	0.033	0.545	0.548

Notes: Standard errors are clustered at the the classroom level. * denotes significance at 0.10; ** at 0.05; *** at 0.01.

Table 5: Students' Intermediate outcomes

	Effort		Motivation and Non-cognitive traits			
	Attendance	Study Hours	Motivation to study hard	Self esteem	Grit	Conscien tiousness
	(1)	(2)	(3)	(4)	(5)	(6)
Panel A						
Standard	0.0235*	-0.970	-0.0712**	-0.0300*	-0.0336	-0.0449
	(0.0130)	(1.036)	(0.0353)	(0.0169)	(0.0233)	(0.0319)
Relative	0.00858	-1.562	-0.0356	-0.0283	-0.0266	-0.0266
	(0.0151)	(1.158)	(0.0388)	(0.0171)	(0.0233)	(0.0340)
R-Squared	0.193	0.076	0.022	0.050	0.049	0.080
Panel B						
Standard	0.0246*	-0.732	-0.0927**	-0.0341*	-0.0336	-0.0527*
	(0.0136)	(1.097)	(0.0379)	(0.0183)	(0.0231)	(0.0313)
Relative	0.00803	-1.304	-0.0478	-0.0261	-0.0153	-0.0120
	(0.0161)	(1.205)	(0.0426)	(0.0192)	(0.0244)	(0.0315)
Standard x Top 15%	-0.00939	-1.214	0.137**	0.0252	-0.000820	0.0471
	(0.0225)	(1.698)	(0.0624)	(0.0350)	(0.0466)	(0.0829)
Relative x Top 15%	-0.00848	-1.826	0.0639	-0.0194	-0.0762	-0.0910
	(0.0226)	(1.988)	(0.0657)	(0.0346)	(0.0470)	(0.0869)
Baseline top 15%	0.0492***	2.988**	-0.0244	0.0348	0.0846**	0.0638
	(0.0165)	(1.469)	(0.0442)	(0.0297)	(0.0335)	(0.0737)
Grade fixed	Yes	Yes	Yes	Yes	Yes	Yes
District fixed	Yes	Yes	Yes	Yes	Yes	Yes
Baseline value	Yes	Yes	Yes	Yes	Yes	Yes
Demographic control	Yes	Yes	Yes	Yes	Yes	Yes
N	7085	5242	5754	5842	5842	5844
R-Squared	0.195	0.078	0.023	0.052	0.052	0.083
Mean of Dep. Var.	0.756	14.526	4.298	2.719	3.259	3.674

Notes: Standard errors are clustered at the the classroom level. * denotes significance at 0.10; ** at 0.05; *** at 0.01.

Table 6: Teachers and Parental response

	Teacher effort index	Parental effort index	Parents mentioned scholar-ship
	(1)	(2)	(3)
Panel A			
Standard	-0.0172	-0.0206	0.126**
	(0.0447)	(0.0455)	(0.0636)
Relative	-0.0265	0.00835	0.0868
	(0.0401)	(0.0439)	(0.0706)
R-Squared	0.091	0.042	0.038
Panel B			
Standard	-0.0199	-0.0203	0.0800
	(0.0461)	(0.0484)	(0.0695)
Relative	-0.0305	0.0217	0.107
	(0.0433)	(0.0467)	(0.0689)
Standard x Top 15%	0.0180	-0.00391	0.276**
	(0.0555)	(0.0571)	(0.111)
Relative x Top 15%	0.0167	-0.0800	-0.0663
	(0.0608)	(0.0548)	(0.132)
Baseline top 15%	0.0189	0.0502	-0.236***
	(0.0432)	(0.0431)	(0.0861)
Grade fixed	Yes	Yes	Yes
District fixed	Yes	Yes	Yes
Baseline value	Yes	Yes	No
Demographic control	Yes	Yes	Yes
N	5838	5778	5848
R-Squared	0.091	0.043	0.042
Mean of Dep. Var.	4.006	4.060	3.409

Notes: Standard errors are clustered at the the classroom level. * denotes significance at 0.10; ** at 0.05; *** at 0.01.

Table 7: Feedback effect: Test Score Impacts

	Final Exam Rank (pct)			Final Exam Score (Norm.)		
	All	Top 15%	Bot 85%	All	Top 15%	Bot 85%
	(1)	(2)	(3)	(4)	(5)	(6)
Panel A						
Feedback	0.850	2.268**	0.822	0.0316	0.0882	0.0299
	(0.596)	(1.117)	(0.721)	(0.0222)	(0.0641)	(0.0260)
R-Squared	0.295	0.254	0.228	0.303	0.249	0.228
Panel B						
Feedback	1.314	2.039	1.539	0.0579	0.137	0.0588
	(1.818)	(2.045)	(2.207)	(0.0666)	(0.0956)	(0.0789)
Standard	-9.027*	-5.212	-7.738	-0.300	-0.190	-0.246
	(5.325)	(5.432)	(4.762)	(0.206)	(0.264)	(0.176)
Relative	-6.278	-2.566	-4.225	-0.188	-0.103	-0.0852
	(5.734)	(4.898)	(5.315)	(0.230)	(0.261)	(0.208)
Standard x FB	0.0188	1.431	-0.408	-0.0282	-0.0322	-0.0335
	(2.026)	(2.873)	(2.450)	(0.0734)	(0.133)	(0.0866)
Relative x FB	-1.021	-0.745	-1.255	-0.0334	-0.0899	-0.0365
	(1.969)	(2.926)	(2.371)	(0.0734)	(0.162)	(0.0858)
Grade fixed	Yes	Yes	Yes	Yes	Yes	Yes
District fixed	Yes	Yes	Yes	Yes	Yes	Yes
Baseline value	Yes	Yes	Yes	Yes	Yes	Yes
Demographic control	Yes	Yes	Yes	Yes	Yes	Yes
N	5188	794	4394	5188	794	4394
R-Squared	0.306	0.263	0.237	0.312	0.255	0.237
Mean of Dep. Var.	51.469	80.275	46.264	-0.180	0.997	-0.393

Notes: Standard errors are clustered at the the classroom level. * denotes significance at 0.10; ** at 0.05; *** at 0.01.

Figure A1: Quiz for program understanding

In TA Chimutu, 3,000 pupils from Standard 5 are participating in this program. They are randomly assigned to Group A, B, and C. All the pupils will be divided into subgroups of 100 pupils in the order of their performance on the previous exam marks. Here are the specifics about each Group:

- Group A: a pupil will receive a present if he/she is ranked at top 15% (450th or above) out of the 3,000 pupils in the final exam.
  .
- Group B: a pupil will receive a present if he/she is ranked at top 15% (15th or above) in his/her subgroup (100 students) in the final exam

- Group C: none of the students in Group C will receive a present.

Sample Question

1. Mary is a Standard 5 student in Singogo Primary School. Her class is assigned to Group C. Is Mary going to receive present?
   a. Yes
   b. No
   c. Not enough information

Quiz

1. Edson is a Standard 5 student in Katete Primary School. His rank in the previous exam was 0.5% (15th out of 3,000) and his class is assigned to Group A. In the final exam, he scored a little lower than before, and was ranked at 7% (238th out of 3,000). Is he going to receive a present?
   a. Yes
   b. No
   c. Not enough information

2. Ethel is a Standard 5 student in Mgona primary school. Her rank in the previous exam was 35% (1,070th out of 3,000), and his class is assigned to **Group B**. So she was included in the subgroup of the students with ranks 1,001st ~ 1,100th. In the final exam, she was ranked at top 20% (600th out of 3,000) and this was top 10% (10th best performance) among her subgroup. Is she going to receive a present?
   a. Yes
   b. No
   c. Not enough information

3. Chikalipo is a Standard 5 student in Chimlamba Primary School. His class is assigned to Group A. In the previous exam, his rank was 64% (1,945th out of 3,000). In which case among below can he receive the present in the final exam?
   a. When he is ranked 63% (1915th out of 3,000)
   b. When he is ranked 0.5% (15th out of 3,000)
   c. He will not receive present

4. Enous is a Standard 5 student in Chang'ana Primary School. His class is assigned to Group B. In the previous exam, his rank was 23% (712th out of 3,000), so he was included in the subgroup of students with ranks between 701st ~ 800th. In which scenario will he receive a present in the final exam? (2 answers)
   a. When he is ranked at 10% (315th out of 3,000) and it was top 13% (13rd best performance) within his subgroup
   b. When he is ranked at 23% (710th out of 3,000) and it was top 10% (10th best performance) within his subgroup
   c. When he is ranked at 23% (710th out of 3,000) and it was top 79% (79th best performance) within his subgroup

5. Angella is a Standard 5 student in Phiri Primary School. Her rank in the previous exam was 83% (2,501st. out of 3,000),. In which group will she have the best chance of receiving a present in the final exam?
   a. Group A
   b. Group B
   c. Group C
   d. He has the same chance in Group A and B

Table A1: Sample Attition

	Dependent Variable: Participated		
	Mid-year Exam	Follow-up Survey	Final Exam
	(1)	(2)	(3)
Panel A			
Standard	0.0197	-0.0157	0.0241
	(0.0170)	(0.0180)	(0.0165)
Relative	0.0104	-0.0269	0.0315*
	(0.0177)	(0.0178)	(0.0160)
N	7085	7085	7085
R-Squared	0.148	0.093	0.085
Mean of Dep. Var.	0.877	0.827	0.897
Panel B			
Feedback		0.00222	-0.00271
		(0.00847)	(0.00667)
R-Squared		0.100	0.101
Panel C			
Feedback		0.00229	-0.00283
		(0.00846)	(0.00669)
Standard		-0.0123	0.0269*
		(0.0171)	(0.0160)
Relative		-0.0213	0.0304*
		(0.0194)	(0.0160)
Grade fixed		Yes	Yes
District fixed		Yes	Yes
Demographic control		Yes	Yes
Futher control		Yes	Yes
N		5832	5832
R-Squared		0.101	0.102
Mean of Dep. Var.		0.837	0.890

Notes: Standard errors are clustered at the the classroom level. * denotes significance at 0.10; ** at 0.05; *** at 0.01.

Table A2: Feedback effect: Students' intermediate outcomes

	Effort		Motivation and Non-cognitive tratis			
	Attendance	Study Hours	Motivation to study hard	Self esteem	Grit	Conscien tiousness
	(1)	(2)	(3)	(4)	(5)	(6)
Panel A						
Feedback	-0.00735	0.452	-0.0168	0.00372	0.00488	0.0169
	(0.00649)	(0.486)	(0.0212)	(0.00947)	(0.0125)	(0.0146)
R-Squared	0.186	0.032	0.015	0.041	0.041	0.063
Panel B						
Feedback	0.00598	0.494	-0.0693*	-0.0363	-0.0413**	0.0356
	(0.00991)	(1.175)	(0.0381)	(0.0285)	(0.0196)	(0.0389)
Standard	0.0374**	-1.274	-0.129**	-0.0441*	-0.0403	-0.0365
	(0.0162)	(1.571)	(0.0531)	(0.0240)	(0.0300)	(0.0412)
Relative	0.0209	-2.128	-0.0490	-0.0625***	-0.0552*	-0.0191
	(0.0183)	(1.634)	(0.0516)	(0.0234)	(0.0306)	(0.0417)
Standard x FB	-0.0174	0.204	0.101*	0.0336	0.0530*	-0.0214
	(0.0149)	(1.405)	(0.0536)	(0.0311)	(0.0294)	(0.0433)
Relative x FB	-0.0150	-0.255	0.0309	0.0617*	0.0585**	-0.0231
	(0.0143)	(1.371)	(0.0491)	(0.0312)	(0.0277)	(0.0454)
Grade fixed	Yes	Yes	Yes	Yes	Yes	Yes
District fixed	Yes	Yes	Yes	Yes	Yes	Yes
Baseline value	Yes	Yes	Yes	Yes	Yes	Yes
Demographic control	Yes	Yes	Yes	Yes	Yes	Yes
N	5832	4352	4813	4866	4866	4868
R-Squared	0.188	0.035	0.017	0.043	0.042	0.064
Mean of Dep. Var.	0.734	13.293	4.271	2.712	3.242	3.651

Notes: Standard errors are clustered at the the classroom level. * denotes significance at 0.10; ** at 0.05; *** at 0.01.

Table A3: Feedback effect: Teachers and Parental response

	Teacher effort index	Parental effort index	Parents mentioned scholar-ship
	(1)	(2)	(3)
Panel A			
Feedback	-0.00434	-0.0220*	-0.0479
	(0.0111)	(0.0117)	(0.0396)
R-Squared	0.089	0.039	0.025
Panel B			
Feedback	-0.0120	-0.0489*	-0.0683
	(0.0227)	(0.0250)	(0.0738)
Standard	-0.0740	-0.0311	0.0782
	(0.0552)	(0.0540)	(0.0871)
Relative	-0.0502	-0.0165	0.0486
	(0.0479)	(0.0536)	(0.0956)
Standard x FB	0.0248	0.0313	0.0692
	(0.0269)	(0.0313)	(0.0956)
Relative x FB	-0.00363	0.0336	-0.0157
	(0.0298)	(0.0309)	(0.0963)
Grade fixed	Yes	Yes	Yes
District fixed	Yes	Yes	Yes
Baseline value	Yes	Yes	No
Demographic control	Yes	Yes	Yes
N	4862	4821	4873
R-Squared	0.091	0.039	0.026
Mean of Dep. Var.	3.998	4.056	3.474

Notes: Standard errors are clustered at the the classroom level. * denotes significance at 0.10; ** at 0.05; *** at 0.01.