I Z A Institute
of Labor Economics
Initiated by Deutsche Post Foundation

DISCUSSION PAPER SERIES

IZA DP No. 13188

When Goal-Setting Forges Ahead but Stops Short

Asad Islam
Sungoh Kwon
Eema Masood
Nishith Prakash
Shwetlena Sabarwal
Deepak Saraswat

APRIL 2020

I Z A Institute
of Labor Economics

Initiated by Deutsche Post Foundation

DISCUSSION PAPER SERIES

IZA DP No. 13188

When Goal-Setting Forges Ahead but Stops Short

Asad Islam
Monash University

Sungoh Kwon
Korea Institute of Public Finance

Eema Masood
World Bank

Nishith Prakash
University of Connecticut, IZA, HiCN and CReAM

Shwetlena Sabarwal
World Bank

Deepak Saraswat
University of Connecticut

APRIL 2020

ISSN: 2365-9793

IZA – Institute of Labor Economics

Schaumburg-Lippe-Straße 5–9
53113 Bonn, Germany

Phone: +49-228-3894-0
Email: publications@iza.org

www.iza.org

ABSTRACT

When Goal-Setting Forges Ahead but Stops Short*

In this study, we use at scale randomized control trial among 18,000 secondary students in 181 schools in Tanzania (Zanzibar) to examine the effects of personal best goal-settings on students' academic performance. We also offer non-financial rewards to students to meet the goals they set. We find that goal-setting has a significant positive impact on student time use, study effort, and self-discipline. However, we do not find any significant impact of goalsetting on test scores. We find that, this could be partially because about 2/3rd of students do not set realistic goals. Third, we find weaker results on time use, study effort, and discipline when we combine goal-setting with non-financial rewards, suggesting that typing goal-setting to extrinsic incentives could weaken its impact. We also find that female students improved on outcomes much more than male students and that students coming from relatively weaker socio-economic backgrounds improved more than their counterparts.

JEL Classification: D9, I20, I25, O15, O55

Keywords: goal-setting, recognition rewards, student performance, Zanzibar

Corresponding author:
Nishith Prakash
365 Fairfield Way, Oak Hall, Room 331
Department of Economics
University of Connecticut
Storrs, CT 06269-1063
USA

E-mail: nishith.prakash@uconn.edu

* We appreciate helpful comments and suggestions from Nathan Fiala, Caroline Hoxby, Tarun Jain, Matt Lowe, Michael Manove, Nirajana Mishra, Dilip Mookherjee, Priya Mukherjee, Abhiroop Mukhopadhyay, Philip Oreopoulos, Daniele Paserman, Gautam Rao, Michael Kremer, Stephen Ross, Frank Schilbach, and seminar and conference participants at Boston University, University of Connecticut, and 14th Annual Conference on Economic Growth and Development, Norms and Behavioral Change Conference (2019), and 17th Midwest International Economic Development Conference. This work was made possible through the leadership and support of Abdulla Mzee Abdulla and Khalid Wazir from the Ministry of Education and Vocation Training (MoEVT) in Zanzibar. We are also grateful for the support of Huma Ali Waheed and Kaboko Nkahinga from the World Bank. Tumainiel Ngowi provided excellent field support. Funding from the Results in Education for All Children (REACH) Trust Fund is gratefully acknowledged. Declarations of Interest: None.

1 Introduction

Despite making tremendous progress in improving school enrollment, many countries struggle with a low level of student learning. For example, in India, among children in grade 8, only 22.1% can do subtraction, and 43.9% can divide (ASER (2017)).[1] Low level of student learning has long-run implications on their lives and represents lost output for the economy as a whole (Michelmore & Dynarski (2017); Hanushek (2009)). Policies aiming to improve academic performance have extensively focused on using financial incentives, such as paying students for their improved test scores.[2] However studies using both randomized and natural experiments on the impact of financial incentives to improve student performance have been mixed and inconclusive.[3] In addition to being expensive and difficult to scale up, evidence from the psychology literature argues that financial incentives might crowd out intrinsic motivation to study (e.g., Cameron & Pierce (1994), and Gneezy *et al.* (2011)). On the other hand, self-selected goals and non-financial incentives, given their self-selected nature may tie well with intrinsic motivation. In this paper, using a large randomized control trial, we test the impact of *goal-setting* and a *non-financial incentive* (recognition award) on *student effort* and *test scores* in Zanzibar.

There is a strong belief in the psychology literature that goals can act as powerful motivators that may affect both thought and action towards improving an outcome (Locke (1968); Locke & Latham (1990); Heath *et al.* (1999)). The foundation of this dates back to the prospect theory that suggests that goals can act as reference points (Kahneman & Tversky (1979)), with the psychological motive of loss aversion causing individuals to want to reach

[1] Similarly, among children in grade VIII, only 13.2% can read grade I level text but not grade II level text, and 72.8% can read grade II level text.

[2] In Fryer (2011), in one of the intervention they paid second graders $2 per book to read and pass a short quiz to confirm they read it.

[3] Fryer (2011) finds that financial incentives had little or no effect on the outcomes for which students received direct incentives, self-reported effort, or intrinsic motivation. In particular, he finds a statistically 0 impact of financial incentives on student achievement. Studies estimating the impact of financial incentives on college students performance have yielded mixed results: Henry *et al.* (2004), Cha & Patel (2010), Scott-Clayton (2011), De Paola *et al.* (2012) and Castleman *et al.* (2014) report positive effects; while Cornwell *et al.* (2005), Angrist *et al.* (2009), Leuven *et al.* (2010), Patel & Rudd (2012) and Cohodes & Goodman (2014) do not find any significant positive effects.

their goals (Heath *et al.* (1999)).[4] Also, goals act as a self-imposed commitment device that is used as a tool to motivate oneself, increase effort, persistence, discipline, self-regulation, etc. Therefore, in the context of education, goal-setting could act as a motivation for students to increase effort to achieve those goals (see Church *et al.* (2001), Locke & Latham (1990), Wiese & Freund (2011)), thereby improving their academic performance. Furthermore, goal-setting can enhance student's interest in the subject matter, increase sensitivity to performance outcomes, prompt self-monitoring of performance attainments, promote student's self-efficacy in learning, and help individuals pursue a level of challenge that optimally exceeds their present capacity.[5]

In this experiment we primarily focus on personal best (PB) goal-setting (goals set by students themselves), as opposed to goals set by others (e.g., teachers, parents, or counselor). PB goal-setting is a goal-setting approach that can be personalized to each student's degree of self-control, therefore encouraging students to focus on personal improvement and striving to outperform their best past personal efforts, rather than the efforts of others or achieving against the absolute criteria of the task (Martin (2006); Elliot (1999); Pintrich (2000); Martin & Elliot (2016a)). This approach has been suggested as one way of optimizing students' academic performance (Martin (2006); Martin (2011)). According to the goal-setting literature, specific and challenging goals leads to better performance since these goals reduces the ambiguity of what is to be achieved (Locke & Latham (2002)). In addition, PB goal-setting is associated positively with growth mindsets, achievement, engagement, and academic outcomes (Burns *et al.* (2017); Martin & Liem (2010); Martin & Elliot (2016b); Martin & Elliot (2016a)).[6] The underlying theory is that by when students choose their own goal it "acts as an internal commitment device meant to overcome problems of self-control" (Royer *et al.*

[4]goal-setting and related motive of loss aversion has also been related to better outcomes in sports (Anderson & Green (2018)).

[5]See Bandura & Schunk (1981), Schunk (1983), Bandura & Cervone (1983), Bandura & Cervone (1986), Bandura (1997), Deci & Ryan (2000) and Csikszentmihalyi (1990) for discussions.

[6]In longitudinal work, Martin & Liem (2010) found that PB goals predict later literacy achievement, numeracy achievement, motivation, and engagement. Similarly, Liem *et al.* (2012) finds that PB goals significantly predict deep learning and academic flow. In research among Chinese students, Yu & Martin (2014) found a positive role for PB goals on students' academic engagement.

(2015); Samek (2016)).[7]

In addition to intrinsic motivation (e.g. goal-setting in our context), economic theory suggests that extrinsic motivations can act as "status incentives" for an agent to increase effort and achieve better outcomes (see Besley & Ghatak (2008)). Ashraf *et al.* (2014) in an experimental study in Zambia finds that employer recognition increases effort and performance. Students, when faced with extrinsic incentives and gains in the form of recognition conditional on an increase in performance, may increase study effort. A combination of an extrinsic and intrinsic incentive (PB goal) may have higher gains than PB goal-setting alone. On the other hand, there is evidence that in cases where the effort is put towards tasks which are *moral* or *social* in nature, the extrinsic incentives may crowd out intrinsic motivations (See Bowles (2008) and Heyman & Ariely (2004)).

Improving the academic performance of students using *cost-effective* and *scalable* incentives has been a challenging endeavor and a focus of researchers studying education. Therefore, from a policy standpoint, PB goal-setting offers a low-cost, scalable option with intrinsic merit beyond its instrumental value in promoting student achievement. In this paper, using a large *at-scale* randomized control trial in *all* secondary schools in Zanzibar (187 schools with 18,281 students in grade 8), we answer two important questions in the field of behavioral development economics. First, do self-set goals provide sufficient impetus for improved student effort and academic performance measures by test scores? Can their efficacy be improved through extrinsic incentives (recognition award) tied directly to goal achievement?

We answer these two questions by randomly assigning the 187 secondary schools into one of the two treatments and one control group. In the first treatment, we encourage students to set their own personal best (PB) goals for improvements in math test scores. In the second treatment, we combine the goal-setting in first treatment with performance-based

[7]goal-setting theory is a key conceptual underpinning of PB goal-setting (Locke & Latham (1990)). According to Martin (2006), Martin (2011), PB goals are closely linked to goal-setting capacity Locke & Latham (2002).

non-financial recognition awards (medals, certificates, backpack, etc.) for achieving the self-set goals. Locke (1968) argues that such incentives affect performance only through their effect on goals. Also, such non-financial recognition awards lead to several non-material benefits, which come in the form of social-recognition from teachers, peers, or society. This recognition is related to the status of the winner of a non-financial award within the group (e.g., the classroom, the school, or society). We test whether these two treatments lead to improved education outcomes of students in secondary schools.

We find that self-set personal goals lead to a significant *positive* impact on self-reported time use, student effort, and self-discipline. However, we find that self-set personal goals *do not* have discernible impacts on student's test scores in the short run. While test scores show a positive and small improvement, it is not statistically different from zero. In particular, we observe a treatment effect in the range of 0.09 to 0.11 of a standard deviation as improvements in time-use, effort, and self-imposed discipline. Most interestingly, we find that treated students allocate more time in studying math (to which the goal-setting is linked) and substitute this increase in time from helping with household work and sleeping. In terms of student effort, we find that improvements in efforts are driven by increased participation in classroom discussions and organized school work. We also look at a measure of student confidence and find that the intervention *did not* have any statistically significant impact. Looking at the treatment effects separately by gender, we find that female students improve on all the main outcomes much more than the male students. Combining the two treatments, i.e., goal-setting with performance-based non-financial recognition award shows similar trends but weaker as compared to the pure goal-setting treatment. We discuss these results in light of the literature on extrinsic versus intrinsic incentives.

We explore various heterogeneities by students' socio-economic backgrounds since they are likely to respond differently to the intervention. We explore this by looking at students with diverse English language skills of parents and by the level of household wealth. Results suggest that students having parents who cannot read and write English improve more on

time use, effort, and self-imposed discipline as compared to students having parents who can read and write English. We observe similar effects when we look at students coming from families with wealth levels below the median of the entire study sample. In particular, students coming from lower wealth levels show larger improvements in time use, effort, and self-imposed discipline as compared to their richer counterparts. In addition to these heterogeneities, we use the distribution and characteristics of the goals set by students to analyze how the impact varies by student's perception of their ability.

This paper contributes to several related literature. Most narrowly, our findings contribute to the literature using experiments to estimate the impacts of self-set personal best goals on academic performance in various settings. Recent experimental studies in the US and Canada use a variety of goal-setting interventions and incentives related to academic performance and find mixed results on academic outputs.[8] Among the closest to our study, Clark *et al.* (2017) in the context of undergraduate students in a public university in the US finds that only the goals which are specific to certain academic tasks show improvements in completion and performance. On the other hand, in developing country context Mukherjee & Poonuganti (2019) find no overall impact of parents involvement in setting goals and aspirations on their kid's academic outcomes in India. Dobriyoni *et al.* (2017) in the context of college education in Canada finds no impact of goal-setting exercises on GPA, course credits, or persistence in subsequent years of education. Lent (2018) using a similar setting finds no impact of goal-setting on undergraduate academic performance and attributes this to the rigidity of set goals. Another related experiment by van Lent & Souverijin (2017) analyzes the effects of setting a goal and increasing its ambitiousness using mentor-student meetings involving first-year university students and finds students in the treatment groups performed better, however, students who were challenged to set a higher goal performed significantly worse than comparable students in the goal treatment. Our paper adds to this growing branch of literature around goal-setting in an academic context where we find evidence that

[8]See Clark *et al.* (2017), Lent (2018), Morisano *et al.* (2010), Levitt *et al.* (2016), and O'Neil *et al.* (1995).

while the outcomes like effort, time-use, and discipline move in the right direction, but for these to translate into improved test scores, we would probably need much larger movements in these behavioral measures to start with.

Additionally, an important contribution of this paper is to extend the goal-setting literature to the context of a developing country and pre-college (secondary school) setting. The targeted student population is of particular interest to the policymakers given very high rates of student drop-out around this age. In Zanzibar, almost half of the students entering secondary schools drop out before the completion. Also, the transition from lower secondary to higher secondary is only 8.4 percent (MOEVT (2017)). Evidence suggests that most students drop-out due to poor performance in lower secondary exit examinations.

We make modest contribution to the very few empirical papers that have analyzed the role of 'status' and 'social recognition' in the context of economics (Ball *et al.* (2001); Markham *et al.* (2002); Charness *et al.* (2010) and Kosfeld & Neckermann (2010)). Although in this paper we do not directly test the pure 'status' dimension of awards and student recognition as predicted by number of theories, we estimate if such awards compliment *(or not)* the impact of PB goal-setting on students' academic performance, especially if tied directly to goal achievement.[9]

Finally, to the best of our knowledge, this is the first paper conducting an *at-scale* randomized experiment related to goal-setting. While in theory smaller-scale experiments can test and inform a potential large scale program rollout, due to governmental and bureaucratic constraints, it does not happen as often. An intervention as cost-effective as goal-setting is easier to roll out at a larger scale hence is better tested at such a large scale. In addition to that, large scale experiments not only circumvent the problem of external validity in a randomized experiment but also avoid the issue of program effects being different at a smaller

[9]There are large number of theories on status and social recognition and predictions in economics (e.g. Ellingsen & Johannesson (2007); Frey (2007); Moldovanu *et al.* (2007); Auriole & Renault (2008); Besley & Ghatak (2008); Dur (2009); Dubey & Geanakopolos (2005); Ederer & Patacconi (2008)). Benabou & Tirole (2006) provide a set of tests for the hypothesis that volunteers are motivated by social-image concerns about their preferences for prosocial behaviors and material rewards.

scale versus at a larger scale (Muralidharan & Niehaus (2017)).

2 Experimental Design

2.1 Context

Zanzibar, off the coast of mainland Tanzania, is a semi-autonomous archipelago, that comprises of two main islands: Unguja and Pemba, and multiple smaller islands around the region. The Government of Zanzibar acts independently from Tanzania on all matters other than foreign policy. Zanzibar's economy is mainly supported by the service industry, with tourism contributing to 51% of the GDP (Mosedale (2010)). The total population of Zanzibar is estimated to be around 1.6 million in 2015 (OCGS (2016)), with around two-thirds living in Unguja. The literacy rate, as defined by the percentage of people above 10 years of age who can read and write simple statements was around 84% in 2016 (MOEVT (2017)). This figure was slightly lower for females at around 79%. Compared to Tanzania as a whole, the literacy rate is 5-10 percentage points higher in Zanzibar (MOEVT (2017)).

Education is considered a basic human right in Zanzibar, and is free at the primary level. The education structure is organized as two years of pre-primary, then six years of Primary schooling starting at six years old. From here students move on to Lower Secondary for a total of four years before starting Advanced secondary school for an additional two years. Once they clear Advanced Secondary, they can move on to Higher Education. The language of instruction is English from Standard 5 onwards, henceforth, all subjects, except Kiswahili, are taught and tested in English.

Student performance in national exams is generally poor. Around one-fifth of all students taking the secondary school entrance exam failed to pass. Students' performance in Mathematics was observed to be especially low. At the lower secondary level, only around half of all students managed to pass the Form 2 exam (lower secondary level or grade 8 and 9), while the rest comprised of those that failed or did not take the exam. High levels of

variation are found across the subjects in the Form 2 exam, with students scoring around a 45% in the Kiswahili on average, while only managing a 15% average in Math. Dropout rates are especially problematic at the ordinary secondary level, with around 30% of the students failing to pass the Form 2 exam, and around half of all students leaving the system before the end of the four-year cycle (MOEVT (2017)).

2.2 Intervention and Timeline

We conduct the nationwide experiment in Zanzibar where all grade eight students in public secondary schools were a part of the study sample. There were a total of 187 secondary schools randomly assigned to two treatments and one control group (see Table 1 for sample sizes). Goals in both treatment arms were set following Martin & Elliot (2016b) and Martin & Elliot (2016a). Treatment announcement was preceded by a baseline data collection and baseline maths and English test. After the treatment was announced and before the endline data collection and test, students were reminded of their goals. Table 2 shows the timeline of the study, interventions, and reminders.

The *Treatment 1* group, also known as "goal-setting" received the personal best goal-setting intervention. In this group, the enumerators introduced the concept of a Personal Best goal to the Form 2 students, using a given script (see Appendix Figure A.1). The enumerators then used an interactive exercise to ensure students understood the meaning behind a personal best goal. Before we asked students to set their goals, we conducted a standardized baseline test for students in English and Mathematics using a curriculum-based assessment specifically developed for the study. Students were asked to set their personal goals soon after the baseline test was completed, and based on their expected score in the baseline test. This was a personal best goal for themselves for a similar exam at the end of the year (about 9 months). Students were asked to think about these goals carefully, allowing for improvement while keeping them realistic.

The *Treatment 2* group, also known as "goal-setting + Recognition" received the personal

best goal-setting intervention as in Treatment 1, but their ability to meet their personal best goals was tied to a Non-Financial recognition reward. These rewards were in the form of certificates of achievement given in a ceremony in front of the whole school. Students were made aware of this reward as part of the given script in Treatment 2 schools (see Appendix Figure A.2).

After the treatment announcements, teachers and head-teachers in the two treatment groups were asked to give students periodic reminders of the goals they had set. Schools also received a poster to display in the school, reminding students about working on their goals every month. Systematic field-based reminders of the Personal Best goal-setting intervention was also undertaken in the two treatment schools. Each student was individually shown the goals they had set for themselves earlier that year as a reminder. Finally students were told that the endline exam would be undertaken at the end of that year, to encourage them to work on their goals.

2.3 Data Collection

Baseline data collection was conducted in February 2016, which included: (i) Survey with the Head Teacher, (ii) Survey for the Form 2 English and Math teachers, and (iii) Form 2 Student Survey and Assessment. At the end of the data collection, the enumerators were instructed to make announcements to the 2 treatment groups on goal-setting exercises, and students in the treatment groups were given a (iv) Treatment Sheet to record their goals. The baseline student sample consists of around 18,281 students from all schools. Endline data collection was conducted in mid to end October 2016, which included: (i) Survey with the Head Teacher, (ii) Survey for the Form 2 English and Math teachers, and (iii) Form 2 Student Survey and Assessment in English and Math. Only students from the baseline were tested in the endline.

2.4 Validity of the Experimental Design

To ensure that the randomization was successful and treatment and control schools were similar before the experiment, we perform a balance test on student and school characteristics respectively in Tables 3 and 4. We do not find any statistically significant difference in school-level characteristics, key demographic characteristics of students, and baseline achievement across treatment and control groups. Most importantly, there are no statistically significant differences across the treatment and control students on age, financial status, and baseline test score.[10] However, there are small imbalances on student gender. About 26% of students were absent during the endline data collection which gives us 13,426 students on which the final analysis is conducted. Table 3 shows that this attrition rate was not statistically different across study groups. After presenting the main results, we will revisit this issue of attrition and attempt to understand and alleviate concerns around its potential impact on the results.

The average age of students in the study is about 16 years and 7 months. Around 55 percent of the students are female; 74 percent reported living with both parents; 6.4 percent are repeating their current grade, and 9.7 percent are new to their respective schools. On average, students reported spending 3 hours a week studying for Mathematics outside of their school, and around 47 percent reported attending exam preparation classes for Mathematics.

2.5 Goal-Setting

Students in both treatment arms set goals in the form of a target score to achieve (out of 20 points) at the endline test. Figure 2 and 3 show the distribution of goals set (out of 20) for both treatment arms. As observed, the majority of students set very high goals. The distribution of set goals is remarkably similar across both the treatment arms, thereby providing evidence against any strategic goal-setting across arms. In an attempt to understand the goal-setting in detail, we plot the distribution of the gap between the set goal and actual

[10] Figure 1 shows a similar distribution of test scores in mathematics across groups at the baseline.

baseline score for both the treatment arms in Figures 4 and 5 respectively. As observed, a majority of students have set very high goals in comparison to the actual baseline performance, and this pattern is similar across both the treatment arms. Most students have aimed at covering a gap of more than 10 points from their baseline score; a gap which is *more than half* of the total points on the exam. Since the goals were set with a reference point of *expected score at baseline test*, this large gap could be a result of the students' *overestimating* their baseline performance, the students being *overambitious* about their future performance or both. We decompose the gap between the goal and the actual baseline score as follows:

$$(Goal - Actual\ \ Baseline\ \ Score) = (Expected\ \ Baseline\ \ Score - Actual\ \ Baseline\ \ Score) + (Goal - Expected\ \ Baseline\ \ Score)$$

Where, the first term on RHS is *overestimation* and second term is *overambition*. We classify a student in any treatment group as *overestimating* their baseline performance if the gap between their expected and actual performance is equal to or more than half of the total points on test i.e. ≥ 10 points. Similarly, a student setting a goal which aims at covering a gap of more than or equal to 10 points from their expected baseline score is termed *overambitious*. Figure 6 and 7 show the distribution of overestimation for both treatment arms, and as observed, the majority of students overestimate their baseline performance by a factor of more than 10 points. Table A-1 shows the proportion of treated students in each of these categories. As observed, 60% of treated students are overestimating their baseline performance while only 6% set overambitious goals.

2.6 Outcome Measures

We analyze the impact of the intervention on *six* key outcome measures. We discuss the construction of these outcome measures below in details:

Student Time-Use: In both the baseline and endline survey, we collected data on time use on

an average weekday on various time use categories. These categories include: *studying and doing homework outside school*, *helping family with household or other type of work*, *sleeping*, *playing games, chatting with friends etc outside school*, *Studying extra for the endline exam*, and *hours studying math outside school*. Responses to these questions in survey are coded on an increasing scale of 1 to 5 with 1 being the lowest and 5 being the highest value.[11] Standardized values of responses to all these questions are converted to a single Anderson's Index (see Anderson (2008)), called *Time-use Index*.[12]

Effort Index: In the endline survey, we collected data on measures of effort students have put in the class and for exams using questions related to their studying habits in the class and for exams. These questions are Likert scale responses to statements like *I studied regularly*, *I tried to do well compared to other students*, *I tried to get a better score than the last year*, *I actively participate in class discussions*, *I prepare and review lessons*, and *I plan and organize my school work*. These statements were ranked by students on a Likert scale of 1 (strongly disagree) to 4 (strongly agree). We combine the standardized values of these responses to form a single Andersons's Index called *Effort Index*.

Self-Discipline Index: We collected student's responses to statements measuring the degree of self-discipline in a student's life. These statements are: *I like to be very good at what I do*, *I can be very disciplined and push myself*, and *I finish whatever I begin*. These statements were ranked by students on a Likert scale of 1 (strongly disagree) to 4 (strongly agree). We combine the standardized values of these responses to form a single Andersons's Index called *Self-Discipline Index*.

Confidence Index: We collected the student's response to statements measuring the level of confidence. These statements are: *I feel very confident in exam*, *I feel very confident when I play with my friends*, and *I feel very confident talking to my teachers and responding to their*

[11] Responses range from *Usually not at all* coded as 1 to *More than X hours* coded as 5.

[12] Responses for *Sleeping*, *Helping with family work*, *Sleeping* and *Playing games etc* are reverse coded as these are likely the substitutes for spending more time in studying.

questions in class. These statements were ranked by students on a Likert scale of 1 (strongly disagree) to 4 (strongly agree). We combine the standardized values of these responses to form a single Andersons's Index called *Confidence Index*.

Aspirations Index: We collected student's responses to statements measuring the level of aspirations. The statements are: *I have high goals and aspirations*, *I do not expect much from my future*, and *I have a desire to pursue further education*.[13] Their statements were ranked by the student on a Likert scale of 1 (strongly disagree) to 4 (strongly agree). We then combine the standardized values of these responses to form a single Anderson's Index called *Aspirations Index*.

Test Score: The goal-setting exercise in both the treatment arms were in connection with the Math test scores. We administered a Math test at baseline followed by the same test (with questions ordered differently) at the endline. We use these endline test scores as our outcome of interest. We standardize the raw scores by creating z-scores for both endline and baseline scores.[14] We also report similar z-scores for the English test which were administered during baseline and endline.

Parent's and Teacher's Efforts Index: In the endline survey we ask students questions related to teacher and parent's effort and we combine them to form indices for teacher and parent's effort.[15] Questions related to teacher's effort are: *Did your teacher assign any homework in last week?*, *Did your teacher give quizzes or tests in last month?*, and *If you had questions or problems, could you discuss with your teacher freely?* Questions related to parent's effort are: *During the last week, have your parents asked about your school life?*, *During the last week, have you worked on school work with your parents?*, and *During the last week, have your parents checked if you did the homework?* We combine the standardized values of these responses to form two Anderson's Index called *Teacher Effort Index* and *Parent Effort Index*.

[13] The statement *I do not expect much from my future* was reverse coded.

[14] We use the control group as the base category. The formula used is: $\frac{\text{(Raw Score - Mean of Control Raw Score)}}{\text{Standard Deviation of Control Raw Score}}$.

[15] The responses to these questions are recorded in Yes or No.

2.7 Estimating Equation

We are interested in estimating the impact of PB goal-setting (Treatment 1: GS) and PB goal-setting with public recognition (Treatment 2: GS + R) on outcomes of interest. We estimate the following equations to evaluate the impact of the two treatments:

$$Y_{is}^{Post} = \beta_0 + \beta_1 T_s^{GS} + \beta_2 T_s^{(GS+R)} + Y_{is}^{Pre} + \epsilon_{is} \tag{1}$$

where, i is the student in school S. Y_{is}^{Post} is the outcome of interest observed at the endline. T_s^{GS} and $T_s^{(GS+R)}$ denotes goal-setting and goal-setting + recognition treatments respectively. Y_{is}^{Pre} is the baseline value of outcome observed at the endline. β_1, and β_2 are our main coefficients of interest and provides the *intent-to-treat* estimate, which is the effect of goal-setting and goal-setting + public recognition on the outcomes of interest. We also estimate a modified version of equation 1 for the pooled treatments ($T_s^{GS} + T_s^{(GS+R)}$). ϵ_{is} is the error term. We cluster the standard errors at the school level since randomization is at the school level.

3 Results

3.1 Average Treatment Effects

We first present the estimates of the impact of goal-setting and goal-setting combined with recognition on our first stage outcomes: *Time-use Index*, *Effort Index*, *Self-Discipline Index*, *Confidence Index*, and *Aspirations Index* in columns 1-5 in Table 5. These are important behavioral changes that have been shown in the literature to be highly predictive of educational outcomes (Heckman *et al.* (2006); Almlund *et al.* (2011); Alan *et al.* (2019)). We present the same for the pooled treatment in Table A-2. We further break down the aggregate indices reported in Table 5 into their individual components and present the estimates in Table 6 to Table 10. This allows us to examine the variables driving the observed effects

in the aggregate index for *Time-use*, *Effort*, *Self-Discipline*, *Confidence*, and *Aspirations*.

We present the estimates of the impact of the two treatment arms for the *Time-Use Index* in column 1 of Table 5 and find that both treatments led to a significant change in student's time use behavior. The effect ranges from 11.3% of a standard deviation (s.d) due to goal-setting alone to 10% of a standard deviation due to goal-setting plus the recognition award. These estimates are statistically significant with a p-value of less than 0.01 and 0.05, respectively. Table A-2 shows the overall effect of goal-setting on the *Time-use Index* by pooling both the treatments (in column 1). The estimates suggest an aggregate effect of 10.7% of an s.d with smaller standard errors. We then present the estimates for individual components of the *Time-use Index* in Table 6 and find important behavioral changes. In particular, we find that the positive impact on the *Time-use Index* is driven by a reduction in helping with household work (column 2 of panel A), reduction in sleeping time (column 3 of panel A), and most importantly through an increase in study time outside school for Math (column 3 of panel B).

In column 2 of table 5, we present the estimates for *Effort Index* and find that while the goal-setting treatment is associated with a 10.6% of a s.d increment in effort in class and for exams, combining goal-setting with recognition generates a positive but smaller and statistically insignificant effect on the *Effort Index*. Pooled results in Table A-2 shows a positive and statistically significant (with a p-value less than 0.1) aggregate effect of goal-setting on *Effort Index* (column 2). We then present the estimates for individual components of the *Effort Index* in Table 7 and find that the improvements in the *Effort Index* is being driven by *effort to get better score than last year* (column 3 of panel A), *active participation in class discussions* (column 1 of panel B), and *planned and organized school work* (column 3 of panel B).

In column 3 of Table 5, we present the estimates for *Self-Discipline Index* and find that the goal-setting only shows a 9% of s.d increment in discipline index, but while the effect of goal-setting and recognition is positive, it is smaller and statistically insignificant.

Pooling both treatments together Table A-2 shows a positive aggregate effect of goal-setting exercise on the *Self-Discipline Index* (column 3). Table 8 shows that all components of the index except the response to statement *"I can be very disciplined and push myself"* show improvements.[16]

In column 4 of Table 5, we analyze if goal-setting affected student's personalities by looking at the impact of the two treatments on a measure of student confidence: *Confidence Index*. With increased time, effort, and discipline students might likely observe an intermediate improvement in their performance (e.g. problems they can solve now which they could not before), and that may result in higher levels of confidence. We find that the impact of the goal-setting treatment on *Confidence Index* is positive but very small and statistically insignificant[17]. Similarly, goal-setting combined with recognition induces a negative but small effect that cannot be statistically distinguished from zero. Column 4 in Table A-2 shows that when both treatments are pooled, there is a null aggregate effect of goal-setting exercise on *Confidence Index*.

In column 5 of Table 5, we analyze if goal-setting affected student's aspirations. We find evidence that the goal-setting intervention had positive but very small and statistically insignificant on *Aspiration Index*, while this is negative and statistically insignificant for goal-setting with recognition arm.[18] In Table A-2, we find that pooling both the treatments has a small negative impact which cannot be distinguished from zero. We find the results on aspirations consistent with the literature in psychology.[19]

So far we have shown that the goal-setting intervention has an impact on various behavioral outcomes, which directly feeds into the students' education production function and *likely* change test scores. We test the impact of the two treatment arms on the z-scores of

[16]Statement "I can be very disciplined and push myself" shows positive impact but is small in magnitude and statistically insignificant.

[17]Table 9 shows that there is no statistically significant impact on components of confidence.

[18]Table 10 shows that there is no statistically significant impact on components of aspiration.

[19]Literature in psychology demonstrates that aspirations are shaped early in a child's life and tend to decline, become less flexible in response to growing understanding of the world (Gutman & Akerman (2008). Among studies that find changes in aspirations among students, it is often a long term intervention like participation in athletics (e.g. see Hwang *et al.* (2016)) that result in these changes.

endline Math test scores in column 5 of Table 5. We find that both the treatments led to a positive but small and statistically insignificant gain in test scores. We present the same for the pooled treatments in Column 5 of table A-2, and find that pooling lowers the standard errors but given the smaller effect size, it remains statistically insignificant. Improving test scores has not been trivial in the education literature and has mostly been concentrated in studies testing expensive interventions which, unlike behavioral interventions, have a direct impact on the cost of getting an education or on classroom instruction.[20] Only a handful of behavioral interventions have shown positive impact on test scores.[21] Our results are consistent with Oreopoulos & Petronijevic (2019) and Dobriyoni *et al.* (2017) who *do not* find the impact of the social psychology interventions on academic performance in their studies in Canada.

Overall, our results suggest that while goal-setting induces behavioral changes in the right direction by increasing students' time use for study, the effort for study, and discipline, however, these effects are probably not large enough to translate to improvements in test scores. Our results also indicate that goal-setting only improved factors having a direct connection to studies and not so much in improvements in personality, such as confidence.

The analysis in Table 5 is looking at five different outcomes for two treatments each (a total of 10 comparisons). Therefore, a conventional statistical significance observed in outcomes does not rule out the presence of "false positives" due to multiple hypothesis testing. We subject all these 10 comparisons to false discovery test as per Benjamin-Hochberg procedure (see Benjamini & Hochberg (1995)) and find that all results which show statistically significant movements pass the B-H test.[22]

[20] Muralidharan *et al.* (2019), Muralidharan & Sundararaman (2011) and Fiala *et al.* (2019) are examples of few such studies.

[21] Few notable examples include Bettinger & Baker (2014) and Alan *et al.* (2019).

[22] With a chosen false discovery rate of 0.1 and 0.2.

3.2 Heterogeneities

In this section we conduct two heterogeneity analyses around the main outcomes by gender, socio-economic status of the households, and estimation of own ability.

3.2.1 Gender

Male and female students might react differently to being in one of the two treatments. Recent studies testing interventions targeted at improving student outcomes either does not explore this possibility, or do not find differential impacts by gender.[23,24]

We analyze the treatment effects separately for each gender and report the results in Panels A and B of Table 11. Panel A suggests that while male students do show positive gains, these are very small and statistically insignificant, except for the time-use. On the other hand, female students in panel B show larger and statistically significant gains in time-use, effort, and self-discipline. They also show a larger magnitude of gain in test score but it is not statistically significant. These results are particularly important since it provides evidence that goal-setting treatment might have appealed more to female students than their male counterparts. We are not aware of any studies in a developing country context that shows the gender differences in such behavioral interventions.

3.2.2 Socio-economic Status

Students belonging to different socio-economic status might demonstrate a varied level of motivation when subjected to the goal-setting treatment. Dobriyoni *et al.* (2017) finds some suggestive evidence that students with English as their mother tongue gained more from goal-setting in the context of college education.[25] Muralidharan *et al.* (2019) finds no differential impact by socio-economic status for an intervention which leads to a substantial change in

[23] Dobriyoni *et al.* (2017) evaluate interventions related to goal-setting in context of college education, but do not explore the effects differentially by gender of students.

[24] Muralidharan *et al.* (2019) do not find differential impact by gender of tutoring intervention by which shows substantial overall effect on test scores of students in Urban India.

[25] These results, however, do not pass multiple hypothesis testing.

test scores.

We analyze this by dividing the sample by: *Parents being able to read and write english* and by *Household wealth*. Panels A and B of Table 12 reports the results for students whose neither parents read/write English and for students whose either parents read/write English. As observed, while both sets of students show gains in time-use, effort, and self-discipline, students with non-english speaking/writing parents show much larger and statistically significant gains as compared to students with parents who can read/write in English. To look at richer versus poorer students, we divide the sample into higher or lower than the median of the asset index at the baseline. We present the estimates in Panels A and B of Table 13. Students from poorer households demonstrate larger and significant gains in time-use, effort, and self-discipline as compared to students from richer households.

Both the set of comparisons by socio-economic status demonstrate that students coming from comparatively disadvantaged backgrounds get larger gains from the goal-setting intervention. While our study is not equipped to delve deeper into the potential reasons, disadvantaged students are likely more motivated to improve or there is more room for improvement for students from the left tail of the distribution in academic performance.

3.2.3 Estimation of Own Ability

In section 2.5 we analyze the distribution of goals set by treated students and find that majority of students have set very high goals, which in large part, is explained by students overestimating their baseline performance. Using the definition of overestimation discussed in section 2.5, we find that students who overestimate their baseline performance scored lower in the baseline math test.[26] Contrary to the literature looking at gender differences in overconfidence, in our sample of treated students we find that girls overestimate their

[26]The difference of 0.19 Z-score is significant with a p-value < 0.001.

baseline performance more than boys.[27,28]

We analyze the treatment effects by dividing the treated sample into students *overestimating* their baseline performance and students *not overestimating* their baseline performance. We then compare these treated samples to the entire control group. We present the estimates in Table 14. We find that students who overestimate their baseline performance (Panel A) do much better in time-use, effort, and self-discipline as compared to students who do not overestimate their baseline performance (Panel B). We do observe some treatment effects on time-use and English scores in Panel B, although the English test was not part of the goal-setting exercise. Since the students who do not overestimate their baseline performance had higher baseline performance than their overestimating counterparts, it cannot be ruled out that the treated groups in Panel B are higher ability on average as compared to control, and the movements in coefficients are capturing underlying baseline differences.

4 Robustness

4.1 Do Teachers and Parents Alter their Behavior?

A natural concern in a cluster level randomization (schools in this study) is that teachers may alter their performance and effort to increase student's performance and in that sense, the treatment effect we observe on certain outcomes may be the result of teachers altering their behavior in connection to the treatment and not of the goal-setting per se. The same concern also holds for parents altering their inputs in children's study. Table A-3 estimates equation 1 with parent's and teachers's effort indices as outcomes and finds that both of them *do not* demonstrate a value statistically different from zero. This analysis attenuates the concern that the observed treatment effects are a result of altered parent's and teacher's

[27]See Dahlbom *et al.* (2011), Croson & Gneezy (2009) and Bengtsson *et al.* (2004) for a review on gender differences in overconfidence.

[28]There are 5% more girls than boys in the overestimating sample. This difference is significant with a p-value < 0.001.

efforts.

4.2 Power of Sample

Retrospectively, we also analyze the statistical power of the sample. Given that this study was on a large scale, spanning all schools in Zanzibar, the power analysis not only tests the strength of the study sample but also provides credible large sample information about movements in outcomes similar to our study. We use the control group distributions of endline outcomes to calculate the minimum detectable effect (MDE) size taking into account the available sample, number of clusters, and the intracluster correlation. As observed in column 2 of table A-4, the MDEs for time-use index, effort index, and discipline index are under 0.1 of an s.d and also well below the observed effect sizes. For the confidence index, while the study is powered to detect a reasonable change of 0.1 s.d, it observes a very small movement of 0.018 of an s.d. On the other hand, for Math z-score, the MDE is much higher than 0.1 which demonstrates the challenge in detecting the change in test scores, largely because test scores are generally highly correlated within clusters (schools in our case). This exercise indicates that studies analyzing changes in test scores need a much larger scale and the number of clusters to account for high intracluster correlation and detect a reasonably low effect size.

4.3 Attrition

As discussed in Section 3.1, we have substantial attrition in the study from baseline to endline. Table A-5 shows that this attrition ranges from 24.07% in (Goals + Recognition) the treatment to 28.25% in the control group. In this sub-section we aim to understand the attrits and if they can potentially induce any *upward bias* in the observed treatment effects.

In Table A-6 we analyze the nature of attrits by looking at the association of attrition with baseline variables. As observed, girls are less likely to attrit compared to boys. Students who are repeating the grade or are new to the school are more likely to attrit. Looking at

time-use and baseline Math test scores, it turns out that attrits had lower scores and fared worse on time-use factors compared to not attrits. Overall, it looks like that the ones who did not participate in endline were worse in baseline academic indicators. However, balance checks in Table 3 show that attrition is not selectively different in treatment vs control and across both treatments.

Any selective attrition across treatment and control is likely to bias the *true* treatment effect upwards if the students who left the control group would (retrospectively) have been selectively *better performers* at endline than treatment group students and/or vice versa. In this case, the means of outcomes in treatment groups would have been lower than currently observed and/or the mean of outcomes in the control group would be higher. This potential difference in means of both groups due to selective attrition would have reduced the statistical power as well as the detectable differences and hence the attrition posits concerns.

To test the gravity of this potential concern, we use the bounding exercise suggested by Lee (2009) and called "Lee-Bounds". In this exercise, we artificially impute the attrits in treatment and control groups selectively in the direction which may have caused an upward bias and measure the required magnitude of selective artificial imputation to render the treatment and control differences statistically insignificant. In our context, we start making treatment selectively weaker by imputing values which generate lower than the observed (non-attrit) treatment mean by a specific factor. Simultaneously, we make the control stronger by imputing values which generate higher than the observed control mean by the same factor.[29] The imputations and conducted along with associated treatment effects until the treatment effect becomes statistically insignificant (p-value > 0.1). Table A-7 shows results from three such imputations. For *Effort Index*, and *Self-Discipline Index*, the imputation of a factor of 0.4 on treatment and control each (i.e. inducing an 80% overall difference) renders the treatment effect to be statistically zero. For the *Time-use Index*, this factor has to 0.55 in treatment and control each (i.e. inducing a 110% overall difference).

[29] We start by a factor of 0.05 each of both treatment and control means and employ increments of 0.05 each in subsequent imputations.

Overall we conclude from this exercise that students who attrit from study groups have to be 80–110% different across treatment and control (more smarter in control versus the treatment and/or vice versa) for there to be an upward bias in the results. Going by the previous discussion in this section, since the attrits performed relatively poorer on academic indicators compared to non-attrits and that the attrition does not seem to be statistically different across both groups, it is unlikely that attrits would have been selectively smarter in control (i.e. selectively worse in treatment group) by a factor of 80–110% at the endline.

5 Discussion and Conclusion

In this paper, we conduct a large scale field experiment in Zanzibar to evaluate the impact of goal-setting on the academic performance of secondary school students. We find sizable and statistically significant impacts on important behavioral outcomes such as time use, effort, and self-discipline which are likely to enter students' education production function. Contrary to the promising results from research from social psychology, we *do not* find an impact on Math test scores in our context. In particular, Morisano *et al.* (2010) find more than half a standard deviation increase in grades for upper-year students at McGill University. Similarly, Schippers *et al.* (2015) finds that goal-setting to significantly reduce inequalities in achievement if implemented early in students' academic careers.[30] However, our findings are consistent with recent experimental studies on *goal-settings* and *nudges* that also *do not* find any impact on test scores (see Dobriyoni *et al.* (2017); Oreopoulos *et al.* (2018); and Oreopoulos & Petronijevic (2019)). Similar to Oreopoulos *et al.* (2018), we find that both treatments led to a significant change in student's time use behavior, but this positive change did not translate into improvements in academic outcomes. An important difference being Oreopoulos *et al.* (2018) was conducted in a developed country. Also, our analysis of goal-setting reveals that a large fraction of students in the treatment schools had set very high goals compared to their baseline performance *(less realistic goals)*, which may

[30] see Morisano *et al.* (2010) for an overview.

have resulted in sub-optimal levels of efforts that do not align well with their ability.

In the second treatment, when we combine goal-setting with a recognition award, we find a weaker impact on outcome measures. We find this result consistent with both theoretical and empirical evidence on extrinsic motivations crowding out intrinsic motivations, in a context in which the utility from outcomes and gains have a stronger moral and/or social component attached to them.[31] Efforts to improve academic performance have a higher degree of morality attached to them compared to efforts towards competitions or at the workplace. Also, receiving social recognition for putting higher efforts towards academic performance may be construed as less moral or less prosocial.[32] Hence, it is plausible that such social comparisons might have diluted the goal-setting *bite* of the intervention.

The heterogeneity analysis suggests that the goal-setting intervention had a slightly larger impact on female students than their male counterparts. Most importantly, we find that the intervention relatively helped students belonging to the weaker socio-economic backgrounds who are likely catching up from lower levels of performance, and have a larger room to improve. Building on our initial observation of students overestimating their baseline performance, we find that students who overestimate their baseline performance demonstrate slightly better performance in outcomes. We also find that low performing students overestimate their baseline scores, which is consistent with the Dunning-Kruger effect.[33] Higher improvements among overestimating students can also be linked to higher risk-taking ability which may result in better innovation (Hershleifier *et al.* (2012)). It can also be explained as a higher marginal improvement from low baseline levels as compared to their high baseline level counterparts.

Overall, the results from this study suggest that while goal-setting seems to move in the

[31] Bowles (2008) show that incentives may be counterproductive and may crowd out intrinsic motivations when incentives may reduce dignity, morality and autonomy.

[32] Heyman & Ariely (2004) shows that efforts in social markets are much less sensitive to compensation than in a monetary market. In a slightly different context, the model by Benabou & Tirole (2006) predicts that as publicity and rewards increase, incentives are more likely to backfire among volunteers whose preference for prosocial activities is most at risk of being misperceived as a preference for rewards.

[33] As described in Kruger & Dunning (1999), the Dunning-Kruger effect is a cognitive bias in which the people with low ability at a task overestimate their ability.

right direction by positively impacting effort, time-use, and self-discipline, the movement in these behavioral outcomes is not enough to have a discernible impact on actual academic performance. This study also highlights the importance of having an accurate idea of own performance/ability to set realistic goals, and being able to achieve those. This study was conducted *at scale* encompassing the entire area of Zanzibar, and the results, therefore, circumvent the issues related to external validity and potential mismatches between trials at a small scale and large scale-ups.

References

Alan, S., Boneva, T., & Ertac, S. 2019. Ever Failed, Try Again, Succeed Better: Results from a Randomized Educational Intervention on Grit. *The Quarterly Journal of Economics*, **134**, 1121–1162.

Almlund, M., Duckworth, A.L., Heckman, J.J., & Kautz, T.D. 2011. Personality Psychology and Economics. *Pages 1–181 of:* Hanushek, E., Machin, S., & Woessman, L. (eds), *Handbook of Economics of Education*. Amsterdam: Elsevier.

Anderson, A., & Green, E.A. 2018. Personal bests as Reference Points. *PNAS*, **115**, 1772–1776.

Anderson, M.J. 2008. Multiple Inference and Gender Differences in the Effects of Early Interventions: A Reevaluation of the Abecedarian, Perry Preschool and Early Training Projects. *Journal of American Statistical Association*, **103**, 1481–1495.

Angrist, J., Lang, D., & Oreopoulos. 2009. Incentives and services for college achievement: evidence from a randomized trial. *Journal of Applied Economics*, **1(1)**, 136–163.

ASER. 2017. *Annual Survey of Education Report (Rural).*

Ashraf, N., Bandiera, O., & Lee, S.S. 2014. Awards unbundled: Evidence from a natural field experiment. *Journal of Economic Behavior and Organization*, **100**, 44–63.

Auriole, E., & Renault, R. 2008. Status and Incentives. *The RAND Journal of Economics*, **39(1)**, 305–326.

Ball, S., Eckel, C.C., Grossman, P.J., & Zame, W. 2001. Status in Markets. *Quarterly Journal of Economics*, **116(1)**, 161–188.

Bandura, A. 1997. *Self-efficacy: The exercise of control.* MacMillan.

Bandura, A., & Cervone, D. 1983. Self-evaluative and self-efficacy mechanisms governing the motivational effects of goal systems. *Journal of personality and social psychology*, **45(5)**, 10–17.

Bandura, A., & Cervone, D. 1986. Differential engagement of self-reactive influences in cognitive motivation. *Organizational behavior and human decision processes*, **38(1)**, 92–113.

Bandura, A., & Schunk, D.H. 1981. Cultivating competence, self-efficacy, and intrinsic interest through proximal self-motivation. *Journal of personality and social psychology*, **41(3)**, 586.

Benabou, R., & Tirole, J. 2006. Incentives and Prosocial Behavior. *American Economic Review*, **96(5)**, 1652–1678.

Bengtsson, C., Persson, M., & Willenhag, P. 2004. Gender and Overconfidence. *Seminar Paper No. 730, Institute for International Economic Studies, Stockholm University.*

Benjamini, Y., & Hochberg, Y. 1995. Controlling for False Discovery Rate: A Practical and Powerful Approach to Multiple Testing. *Journal of Royal Statistical Society, Series B (Methodological)*, **57(1)**, 289–300.

Besley, T., & Ghatak, M. 2008. Status Incentives. *American Economic Review*, **98(2)**, 206–211.

Bettinger, E.P., & Baker, R.B. 2014. The Effects of Student Coaching: An Evaluation of a Randomized Experiment in Student Advising. *Educational Evaluation and Policy Analysis*, **36**, 3–19.

Bowles, S. 2008. Policies Designed for Self-Interested Citizens May Undermine "The Moral Sentiments": Evidence from Economic Experiments. *Science*, **320(5883)**, 1605–1609.

Burns, M.K., Cook, C.R., & Kilgus, S. 2017. Advancing the science and practice of precision education to enhance student outcomes. *Journal of Social Psychology*, **66**.

Cameron, J., & Pierce, W.D. 1994. Reinforcement, Reward, and Intrinsic Motivation: A Meta-Analysis. *Review of Educational Research*, **64(3)**, 363–423.

Castleman, B., Terry Long, B., Avery, C., Murnane, R., & Goodman, J. 2014. The Impact of Partial and Full Merit Scholarships on College Entry and Success: Evidence from the Florida Bright Futures Scholarship Program. *Working Paper*.

Cha, P., & Patel, R. 2010. Rewarding Progress, Reducing Debt: Early Results from the Performance-Based Scholarship Demonstration in Ohio. *MDRC*.

Charness, G., Masclet, D., & Claire Villeval, M. 2010. Competitive Preferences and Status as an Incentive: Experimental Evidence. *IZA DP No. 5034*.

Church, M.A., Elliot, A.J., & Gable, S.L. 2001. Perceptions of classroom environment, achievement goals, and achievement outcomes. *Journal of Educational Psychology*, **93(1)**, 43–54.

Clark, D., Gill, D., Prowse, V., & Rush, M. 2017. Using Goals to Motivate College Students: Theory and Evidence from Field Experiments. *NBER Working Paper No. 23638*.

Cohodes, S.R., & Goodman, J.S. 2014. Merit Aid, College Quality, and College Completion: Massachusetts' Adams Scholarship as an In-Kind Subsidy. *American Economic Journal: Applied Economics*, **6(4)**, 251–285.

Cornwell, C.M., Hee Lee, K., & Mustard, D.B. 2005. Student responses to merit scholarship retention rules. *Journal of Human Resources*, **XL(4)**, 895–917.

Croson, R., & Gneezy, U. 2009. Gender Differences in Preferences. *Journal of Economic Literature*, **47(2)**, 448–74.

Csikszentmihalyi, M. 1990. Literacy and intrinsic motivation. *Daedalus*, 115–140.

Dahlbom, L., Jakobsson, A., Jakobsson, N., & Kotsadam, A. 2011. Gender and overconfidence: are girls really overconfident? *Applied Economics Letter*, **18(4)**, 325–327.

De Paola, M., Scoppa, V., & Nistico, R. 2012. Monetary incentives and student achievement in a depressed labor market: results from a randomized experiment. *Journal Hum. Cap.*, **6(1)**, 56–85.

Deci, E.L., & Ryan, R.M. 2000. The "what" and "why" of goal pursuits: Human needs and the self-determination of behavior. *Psychological Inquiry*, **11(4)**, 227–268.

Dobriyoni, C.R., Oreopoulos, P., & Petronijevic, U. 2017. Goal Setting, Academic Reminders, and College Success: A large-scale Field Experiment. *NBER Working Paper No. 23738*.

Dubey, P.K., & Geanakopolos, J. 2005. Grading in Games of Status: Marking Exams and Setting Wages. *Cowles Foundation Discussion Paper No. 1544R, Yale University*.

Dur, R. 2009. Gift Exchange in the Workplace: Money or Attention? *Journal of European Economic Association*, **7(2-3)**, 550–560.

Ederer, F., & Patacconi, A. 2008. *Mimio, MIT and Oxford*.

Ellingsen, T., & Johannesson, M. 2007. Paying Respect. *Journal of Economic Perspectives*, **21(4)**, 135–150.

Elliot, A. J. 1999. Approach and avoidance motivation and achievement goals. *Educational Psychologist*, **34**, 169–189.

Fiala, N., Garcia-Hernandez, A., Narula, K., & Prakash, N. 2019. Wheels of Change: Impact of Bicycles on Female Education and Empowerement in Zambia. *Working Paper*.

Frey, B.S. 2007. Awards as Compensation. *European Management Review*, **4(1)**, 6–14.

Fryer, R. 2011. Financial Incentives and Student Achievement: Evidence from Randomized Trials. *Quarterly Journal of Economics*, **126(4)**, 1755–1798.

Gneezy, U., Meier, S., & Rey-Biel, P. 2011. When and Why Incentives (Don't) Work to Modify Behavior. *Journal of Economic Perspectives*, **25**, 191–210.

Gutman, L.M., & Akerman, R. 2008. *Determinants of Aspirations*.

Hanushek, E.A. 2009. The Economic Value of Education and Cognitive Skills. *Pages 39–56 of:* Sykes, G., Schneider, B., & Plank, D.N. (eds), *Handbook of Education Policy Research*. New York: NY: Routledge.

Heath, C., Larrick, R., & Wu, G. 1999. Goals as Reference Points. *Cognitive Psychology*, **38(1)**, 79–109.

Heckman, J.J., Stixrud, J., & Urzua, S. 2006. The Effects of Cognitive and Noncognitive Abilities on Labor Market Outcomes and Social Behavior. *Journal of Labor Economics*, **24(3)**, 411–482.

Henry, G.T., Rubenstein, Ross, & Bugler, Daniel T. 2004. Is HOPE enough? Impacts of receiving and losing merit-based financial aid. *Education Policy*, **18(5)**, 686–709.

Hershleifier, D., Low, A., & Teoh, S.H. 2012. Are Overconfident CEOs Better Innovators? *The Journal of Finance*, **67(4)**, 1457–1498.

Heyman, J.E., & Ariely, D. 2004. Effort for Payment. *Psychological Science*, **15(11)**, 787–793.

Hwang, S., Feltz, D.L., Kietzmann, L.A., & Diemer, M.A. 2016. Sport Involvement and educational outcomes of high school students: A longitudinal study. *Youth and Society*, **48(6)**, 763–785.

Kahneman, D., & Tversky, A. 1979. Prospect Theory: An Analysis of Decision under Risk. *Econometrica*, **47(2)**, 263–292.

Kosfeld, M., & Neckermann, S. 2010. Getting More Work for Nothing? Symbolic Awards and Worker Performance. *IZA DP No. 5040*.

Kruger, J., & 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*, **77(6)**, 1121–1134.

Lee, D.S. 2009. Training, Wages, and Sample Selection: Estimating Sharp Bounds on Treatment Effects. *Review of Economic Studies*, **76**, 1071–1102.

Lent, M.V. 2018. Goal Setting under Uncertainity: A Field Experiment on Rigid and Flexible Goals. *Working Paper*.

Leuven, E., Oosterbeek, H., & van der Klaauw, B. 2010. The effect of financial rewards on students achievement: Evidence from a randomized experiment. *Journal of European Economic Association*.

Levitt, S.D., List, J.A., Neckermann, S., & Sadoff, S. 2016. The Behavioralist Goes to School: Leveraging Behavioral Economics to Improve Educational Performance. *American Economic Journal: Economic Policy*, **8(4)**, 183–219.

Liem, G.A.D., Ginns, P., & Martin, A.J. 2012. Personal best goals and academic and social functioning: A longitudinal perspective. *Learning and Instruction*, **22(3)**, 222–230.

Locke, E.A. 1968. Toward a Theory of Task Motivation and Incentives. *Organizational Behavior and Human Performance*, **3(2)**, 157–189.

Locke, E.A., & Latham, G.P. 1990. *A Theory of Goal Setting and Task Performance*. Englewood Cliffs, NJ, US: Prentice-Hall, Inc.

Locke, E.A., & Latham, G.P. 2002. Building practically useful theory of goal setting and task motivation. *American Psychologist*, **57**, 705–712.

Markham, S.E., Dow Cott, K., & McKee, G.H. 2002. Recognizing Good Attendance: A Longitudinal, Quasi-Experimental Field Study. *Personnel Psychology*, **55(3)**, 639–660.

Martin, A. J. 2006. Personal Bests (PBs): A proposed multidimensional model and empirical analysis. *British Journal of Educational Psychology*, **76**, 803–825.

Martin, A.J. 2011. Personal best (PB) approaches to academic development: Implications for motivation and assessment. *Educational Practice and Theory*, **33**, 93–99.

Martin, A.J., & Elliot, A.J. 2016a. The role of personal best (PB) and dichotomous achievement goals in students academic motivation and engagement: A longitudinal investigation. *Educational Psychology*, **36(7)**, 1285–1302.

Martin, A.J., & Elliot, A.J. 2016b. The role of personal best (PB) goal setting in students' academic achievement gains. *Learning and Individual Differences*, **45**, 222–227.

Martin, A.J., & Liem, G.A.D. 2010. Academic personal bests (PBs), engagement, and achievement: A cross-lagged panel analysis. *Learning and Individual Differences*, **20(3)**, 265–270.

Michelmore, K., & Dynarski, S. 2017. The Gap within the Gap: Using Longitudinal Data to Understand Income Differences in Educational Outcomes. *AERA Open*, **3**, 1–18.

MOEVT. 2017. *Zanzibar Education Development Plan II 2017/18 – 2021/22*.

Moldovanu, B., Sela, A., & Shi, X. 2007. Contests for Status. *Journal of Political Economy*, **115(2)**, 338–363.

Morisano, D., Hirsh, J.B., Peterson, J.B., Pihl, R.O., & Shore, B.M. 2010. Settling, Elaborating and Reflecting on Personal Goals Improves Academic Performance. *Journal of Applied Psychology*, **95(2)**, 255–264.

Mosedale, J. 2010. *Political Economy and Tourism: A critical perspective*. Routledge.

Mukherjee, P., & Poonuganti, S. 2019. Engaging Parents in Learning: Experimental Evidence from Rural India. *Working Paper*.

Muralidharan, K., & Niehaus, P. 2017. Experimentation at Scale. *Journal of Economic Perspectives*, **31**, 103–124.

Muralidharan, K., & Sundararaman, V. 2011. Teacher Performance Pay: Experimental Evidence from India. *Journal of Political Economy*, **119(1)**, 39–77.

Muralidharan, K., Singh, A., & Ganimian, A.J. 2019. Disrupting Education? Experimental Evidence on Technology-Aided Instruction in India. *American Economic Review*, **109**, 1426–60.

OCGS. 2016. *Zanzibar Household Budget Survey (2014-15)*.

O'Neil, H.F.Jr., Sugrue, B., & Baker, E.L. 1995. Effects of Motivational Interventions on the National Assessment of Educational Progress Mathematics Performance. *Educational Assessment*, **3(2)**, 135–157.

Oreopoulos, P., & Petronijevic, U. 2019. The Remarkable Unresponsiveness of College Students to Nudging And What We Can Learn from It. *NBER Working Paper No. 26059*.

Oreopoulos, P., Patterson, R.W., Petronijevic, U., & Pope, N.G. 2018, DOI =. When Studying and Nudging Don't Go as Planned: Unsuccessful Attempts to Help Traditional and Online College Students. *NBER Working Paper No. 25036*.

Patel, R., & Rudd, T. 2012. Can Scholarships Alone Help Students Succeed? *Working Paper*.

Pintrich, P.R. 2000. Multiple goals, multiple pathways: The role of goal orientation in learning and achievement. *Journal of Educational Psychology*, **92**, 544–555.

Royer, H., Stehr, M., & Sydnor, J. 2015. Incentives, Commitments, and Habit Formation in Exercise: Evidence from a Field Experiment with Workers at a Fortune-500 Company. *American Economic Journal: Applied Economics*, **7(3)**, 51–84.

Samek, A. 2016. Gifts and Goals: Behavioral Nudges to Improve Child Food Choice at School. *CESR-Schaeffer Working Paper No. 2016-007*.

Schippers, M.C., West, M.A., & Dawson, J.F. 2015. Team Reflexivity and Innovation: The Moderating Role of Team Context. *Journal of Management*, **41(3)**, 769–788.

Schunk, D.H. 1983. Developing children's self-efficacy and skills: The roles of social comparative information and goal setting. *Contemporary Educational Psychology*, **8(1)**, 76–86.

Scott-Clayton, J. 2011. On money and motivation: a quasi experimental analysis of financial incentives for college achievement. *Journal of Human Resources*, **46(3)**, 614–646.

van Lent, M., & Souverijin, M. 2017. Goal Setting and Raising the Bar: A Field Experiment. *Working Paper*.

Wiese, B.S., & Freund, A.M. 2011. Goal progress makes one happy, or does it? Longitudinal findings from the work domain. *Journal of Occupational and Organizationall Psychology*, **78(2)**, 287–308.

Yu, K., & Martin, A.J. 2014. Personal best (PB) and 'classic' achievement goals in the Chinese context: Their role in predicting academic motivation, engagement and buoyancy. *Educational Psychology*, **34(5)**, 635–658.

Tables

Table 1: Sample Size at Baseline

	(1)	(2)
Study Group	No. of Schools	No. of Students
Control	62	7,105
Goal-Setting	64	5,962
Goal-Setting + Recognition	61	5,214
Total	187	18,281

Notes: This table reports the baseline sample size (both number of schools and number of students) for each of the study groups.

Table 2: Study Timeline

Month/Year	Activities
January, 2016	Randomization and Designing Instruments
February, 2016	Baseline Data Collection + Baseline Tests + goal-setting
August, 2016	Goal Reminders to Students
October, 2016	Endline Data Collection + Endline Tests

Notes: This table shows the timeline of field activities, data collection and rollout of interventions.

Table 3: Balance on Student Characteristics

	Mean (SD)				P-Value		
	(1)	(2)	(3)	(4)	(5)	(6)	(7)
	Overall	Control	GS	GS + R	GS vs. Control	GS + R vs. Control	GS vs. GS + R
Male	0.448	0.457	0.412	0.476	0.035	0.326	0.024
	(0.004)	(0.006)	(0.006)	(0.007)			
Age	16.603	16.612	16.555	16.643	0.48	0.68	0.599
	(0.010)	(0.015)	(0.017)	(0.017)			
Is the student repeating the current grade?	0.064	0.071	0.053	0.068	0.062	0.811	0.139
	(0.002)	(0.003)	(0.003)	(0.004)			
Whether father can read and write in English	0.679	0.683	0.706	0.641	0.364	0.16	0.101
	(0.003)	(0.006)	(0.006)	(0.007)			
Whether mother can read and write in English	0.53	0.532	0.548	0.507	0.592	0.485	0.525
	(0.004)	(0.006)	(0.006)	(0.007)			
Household asset index	0	-0.007	0.065	-0.065	0.428	0.543	0.383
	(0.007)	(0.011)	(0.010)	(0.013)			
Baseline English test z-score	0.043	0	0.121	0.011	0.182	0.911	0.377
	(0.008)	(0.012)	(0.014)	(0.015)			
Baseline Math test z-score	0.047	0	0.096	0.054	0.374	0.631	0.672
	(0.008)	(0.012)	(0.013)	(0.015)			
Spend more than 30 minutes in Math (baseline)	0.661	0.66	0.674	0.648	0.501	0.605	0.468
	(0.004)	(0.006)	(0.006)	(0.007)			
Absent at the endline exam	0.266	0.282	0.267	0.241	0.614	0.169	0.3203
	(0.003)	(0.005)	(0.006)	(0.006)			
Observations	18,281	7,105	5,962	5,214			

Notes: This table reports the balance test for various student level variables captured in baseline survey. Means, standard deviations and p-value of differences is reported by GS (Goal-Setting), GS + R (Goal-Setting + Recognition) and Control group. Standard errors are clustered at the level of school.

Table 4: Balance on School Characteristics

	Mean (SD)				P-Value		
	(1)	(2)	(3)	(4)	(5)	(6)	(7)
	Overall	Control	GS	GS + R	GS vs. Control	GS + R vs. Control	GS vs. GS + R
Total students in F2	132.595	143.86	132.964	120.768	0.699	0.289	0.631
	(10.229)	(17.626)	(21.954)	(12.703)			
Total qualified teachers in F2	4.832	4.638	4.946	4.951	0.545	0.412	0.992
	(0.184)	(0.262)	(0.435)	(0.275)			
Student-teacher ratio in F2	28.228	29.151	28.054	27.341	0.812	0.626	0.876
	(1.696)	(2.646)	(3.767)	(2.592)			
Does this school have two shifts?	0.602	0.583	0.667	0.553	0.463	0.793	0.311
	(0.046)	(0.083)	(0.076)	(0.082)			
Form 2 pass rate in 2015 for English	50.99	50.108	51.582	51.351	0.861	0.88	0.978
	(3.342)	(5.724)	(6.088)	(5.849)			
Form 2 pass rate in 2015 for Math	44.153	38.796	47.474	47.194	0.33	0.309	0.974
	(3.477)	(6.023)	(6.445)	(5.541)			
Form 2 pass rate in 2015 for Science	48.731	46.033	47.703	53.597	0.85	0.329	0.504
	(3.411)	(5.43)	(6.901)	(5.428)			
Average teaching experience in month	150.491	139.479	149.542	162.742	0.541	0.19	0.397
	(6.724)	(13.055)	(9.991)	(11.91)			
Observations	187	62	64	61			

Notes: This table reports the balance test for various school level variables captured in baseline survey. Means, standard deviations and p-value of differences is reported by GS (Goal-Setting), GS + R (Goal-Setting + Recognition) and Control group.

Table 5: Main Results

	(1)	(2)	(3)	(4)	(5)	(6)	(7)
Dependent Variable:	Time-use Index	Effort Index	Self-discipline Index	Confidence Index	Aspirations Index	Math Z-score	English Z-score
Goal-Setting	0.113***	0.106**	0.090**	0.018	0.018	0.056	0.062
	(0.042)	(0.048)	(0.043)	(0.034)	(0.039)	(0.071)	(0.065)
Goal-Setting + Recognition	0.100**	0.051	0.069	-0.010	-0.026	0.026	0.065
	(0.044)	(0.055)	(0.044)	(0.043)	(0.040)	(0.068)	(0.059)
Observations	12,715	11,908	13,049	12,981	12,145	13,426	13,426
Baseline Outcome Controlled	Yes	No	No	No	Yes	Yes	Yes
B-H Passed (Goal-Setting)	Yes	Yes	Yes	N/A	N/A	N/A	N/A
P-Value (Goal-Setting)	(0.009)	(0.027)	(0.035)	-	-	-	-
B-H Passed (Goal-Setting + Recognition)	Yes	N/A	N/A	N/A	N/A	N/A	N/A
P-Value (Goal-Setting + Recognition)	(0.025)	-	-	-	-	-	-

Notes: This table reports the impact of interventions of key outcomes of interest: Time-use Index, Effort Index, Self-Discipline Index, Confidence Index, Aspirations Index, Math test score, and English test score. Construction of these indices is discussed in Section 2. All the results are subjected to Benjamin-Hochberg correction and last set of rows of the table reports if they pass the correction criteria (P-values in parenthesis). Standard errors are clustered at the level of school. *p<0.1; **p<0.05; ***p<0.01.

Table 6: Components of Time-use Index

	Panel A		
	(1)	(2)	(3)
Dependent Variable:	Studying/Homework Outside School	Household work	Sleeping
Goal-Setting	0.042	-0.056	-0.087**
	(0.040)	(0.038)	(0.038)
Goal-Setting + Recognition	0.014	-0.106***	-0.039
	(0.043)	(0.040)	(0.041)
Observations	13,273	13,250	13,219
Baseline Outcome Controlled	Yes	Yes	Yes
	Panel B		
	(1)	(2)	(3)
Dependent Variable:	Games/Leisure time Outside school	Studying extra for Endline exam	Studying Math Outside school
Goal-Setting	-0.043	0.022	0.067*
	(0.027)	(0.048)	(0.039)
Goal-Setting + Recognition	0.004	0.023	0.059
	(0.031)	(0.059)	(0.047)
Observations	13,258	13,310	13,241
Baseline Outcome Controlled	Yes	Yes	Yes

Notes: This table reports the result of estimating equation 1 on individual components of Time-use Index. All the components are discussed in Section 2. Standard errors are clustered at the level of school. *p<0.1; **p<0.05; ***p<0.01.

Table 7: Components of Effort Index

	Panel A		
	(1)	(2)	(3)
Dependent Variable:	Studied regularly	Tried to do well	Tried to get better score
Goal-Setting	0.055	0.058	0.069*
	(0.038)	(0.040)	(0.039)
Goal-Setting + Recognition	0.033	0.032	0.050
	(0.040)	(0.043)	(0.044)
Observations	12,418	12,903	13,028
Baseline Outcome Controlled	No	No	No
	Panel B		
	(1)	(2)	(3)
Dependent Variable:	Participated in class discussions	Prepared Lessons	Organized school work
Goal-Setting	0.076**	0.046	0.056*
	(0.034)	(0.034)	(0.032)
Goal-Setting + Recognition	0.056	0.000	0.012
	(0.042)	(0.042)	(0.034)
Observations	12,601	11,902	12,490
Baseline Outcome Controlled	No	No	No

Notes: This table reports the result of estimating equation 1 on individual components of Effort Index. All the components are discussed in Section 2. Standard errors are clustered at the level of school. *p<0.1; **p<0.05; ***p<0.01.

Table 8: Components of Self-Discipline Index

	(1)	(2)	(3)
Dependent Variable:	Be good at what I do	Be disciplined and push myself	Finish whatever I begin
Goal-Setting	0.078*	0.025	0.077**
	(0.043)	(0.032)	(0.036)
Goal-Setting + Recognition	0.074*	0.011	0.052
	(0.041)	(0.037)	(0.035)
Observations	13,343	13,339	13,152
Baseline Outcome Controlled	No	No	No

Notes: This table reports the result of estimating equation 1 on individual components of Self-Discipline Index. All the components are discussed in Section 2. Standard errors are clustered at the level of school. *p<0.1; **p<0.05; ***p<0.01.

Table 9: Components of Confidence Index

	(1)	(2)	(3)
Dependent Variable:	Feel confident in Exam	Feel confident with friends	Feel confident when interacting with teachers
Goal-Setting	0.016	-0.024	0.045
	(0.031)	(0.034)	(0.029)
Goal-Setting + Recognition	-0.024	-0.032	0.031
	(0.039)	(0.030)	(0.037)
Observations	13,177	13,169	13,392
Baseline Outcome Controlled	No	No	No

Notes: This table reports the result of estimating equation 1 on individual components of Confidence Index. All the components are discussed in Section 2. Standard errors are clustered at the level of school. *p<0.1; **p<0.05; ***p<0.01.

Table 10: Components of Aspiration Index

	(1)	(2)	(3)
Dependent Variable:	Having high goals and expectations	Do not expect much in future	Desires further Education
Goal-Setting	0.042	0.016	0.000
	(0.033)	(0.043)	(0.039)
Goal-Setting + Recognition	0.015	0.043	-0.034
	(0.037)	(0.046)	(0.044)
Observations	12,763	12,485	13,340
Baseline Outcome Controlled	Yes	Yes	Yes

Notes: This table reports the result of estimating equation 1 on individual components of Aspiration Index. All the components are discussed in Section 2. Standard errors are clustered at the level of school. *p<0.1; **p<0.05; ***p<0.01.

Table 11: Heterogeneous Effects - By Gender

	(1)	(2)	(3)	(4)	(5)	(6)	(7)
Dependent Variable:	Time-use Index	Effort Index	Self-discipline Index	Confidence Index	Aspiration Index	Math Z-score	English Z-score
				Panel A: Male Students			
Goal-Setting	0.079*	0.089	0.053	0.051	0.010	0.010	0.045
	(0.047)	(0.062)	(0.049)	(0.047)	(0.047)	(0.087)	(0.069)
Goal-Setting + Recognition	0.071	0.055	0.020	0.031	-0.028	0.009	0.037
	(0.055)	(0.067)	(0.052)	(0.050)	(0.052)	(0.096)	(0.067)
Observations	5,132	4,854	5,315	5,305	4,930	5,454	5,454
Baseline Outcome Controlled	Yes	No	No	No	Yes	Yes	Yes
				Panel B: Female Students			
Goal-Setting	0.130***	0.107**	0.115**	-0.010	0.020	0.102	0.089
	(0.049)	(0.054)	(0.051)	(0.038)	(0.047)	(0.070)	(0.070)
Goal-Setting + Recognition	0.121**	0.046	0.106**	-0.041	-0.025	0.041	0.085
	(0.051)	(0.059)	(0.048)	(0.049)	(0.045)	(0.053)	(0.059)
Observations	7,583	7,054	7,734	7,676	7,215	7,972	7,972
Baseline Outcome Controlled	Yes	No	No	No	Yes	Yes	Yes

Notes: This table reports the heterogeneity (by gender) of the impact of interventions on main outcomes of interest: Time-use Index, Effort Index, Self-Discipline Index, Confidence Index, Aspiration Index, Math test score, and English test score. Construction of these indices is discussed in Section 2. Panel A has the sample of only male students while Panel B reports the results for the sample of female students. Standard errors are clustered at the level of school. *p<0.1; **p<0.05; ***p<0.01.

Table 12: Heterogeneous Effects - By Parents English Language skill

	(1)	(2)	(3)	(4)	(5)	(6)	(7)
Dependent Variable:	Time-use Index	Effort Index	Self-discipline Index	Confidence Index	Aspiration Index	Math Z-score	English Z-score
			Panel A: Neither parent can read/write English				
Goal-Setting	0.232***	0.136*	0.121*	0.007	0.077	0.038	0.065
	(0.060)	(0.074)	(0.067)	(0.057)	(0.058)	(0.066)	(0.067)
Goal-Setting + Recognition	0.136**	0.079	0.064	-0.012	-0.023	-0.002	0.059
	(0.057)	(0.087)	(0.067)	(0.060)	(0.063)	(0.063)	(0.060)
Observations	2,940	2,737	3,036	3,026	2,782	3,150	3,150
Baseline Outcome Controlled	Yes	No	No	No	Yes	Yes	Yes
			Panel B: Either parent can read/write English				
Goal-Setting	0.079*	0.097**	0.081*	0.020	0.000	0.061	0.059
	(0.045)	(0.048)	(0.042)	(0.036)	(0.041)	(0.076)	(0.069)
Goal-Setting + Recognition	0.097**	0.046	0.074*	-0.006	-0.022	0.039	0.075
	(0.049)	(0.054)	(0.044)	(0.048)	(0.039)	(0.075)	(0.063)
Observations	9,775	9,171	10,013	9,955	9,363	10,276	10,276
Baseline Outcome Controlled	Yes	No	No	No	Yes	Yes	Yes

Notes: This table reports the heterogeneity (by english language skill of Parents) of the impact of interventions on main outcomes of interest: Time-use Index, Effort Index, Self-Discipline Index, Confidence Index, Aspiration Index, Math test score, and English test score. Construction of these indices is discussed in Section 2. Panel A reports the results for the sample where neither parents can read/write in English. Panel B reports the results for the sample where either parents (at least one) can read/write in English. Standard errors are clustered at the level of school. *p<0.1; **p<0.05; ***p<0.01.

Table 13: Heterogeneous Effects - By Wealth

	(1)	(2)	(3)	(4)	(5)	(6)	(7)
Dependent Variable:	Time-use Index	Effort Index	Self-discipline Index	Confidence Index	Aspiration Index	Math Z-score	English Z-score
				Panel A: Below Median Wealth at Baseline			
Goal-Setting	0.151***	0.115*	0.123**	-0.025	-0.034	0.012	0.054
	(0.049)	(0.065)	(0.054)	(0.045)	(0.049)	(0.063)	(0.059)
Goal-Setting + Recognition	0.132***	0.064	0.099*	0.007	-0.039	-0.013	0.064
	(0.044)	(0.071)	(0.055)	(0.049)	(0.046)	(0.061)	(0.052)
Observations	6,451	5,984	6,632	6,586	6,118	6,852	6,852
Baseline Outcome Controlled	Yes	No	No	No	Yes	Yes	Yes
				Panel B: Above Median Wealth at Baseline			
Goal-Setting	0.071	0.098**	0.056	0.052	0.056	0.089	0.048
	(0.049)	(0.045)	(0.044)	(0.040)	(0.045)	(0.087)	(0.078)
Goal-Setting + Recognition	0.068	0.036	0.038	-0.030	-0.009	0.072	0.074
	(0.062)	(0.052)	(0.047)	(0.056)	(0.053)	(0.090)	(0.072)
Observations	6,264	5,924	6,417	6,395	6,027	6,574	6,574
Baseline Outcome Controlled	Yes	No	No	No	Yes	Yes	Yes

Notes: This table reports the heterogeneity (by baseline wealth level of student's household) of the impact of interventions on main outcomes of interest: Time-use Index, Effort Index, Self-Discipline Index, Confidence Index, Aspiration Index, Math test score, and English test score. Construction of these indices is discussed in Section 2. Panel A reports the results for sample which is below the median level of wealth at baseline and Panel B reports the results for sample which is above the median level of wealth at baseline. Wealth is measured using the standardized index of the sum of all the self-reported assets owned by the household. Standard errors are clustered at the level of school. *p<0.1; **p<0.05; ***p<0.01.

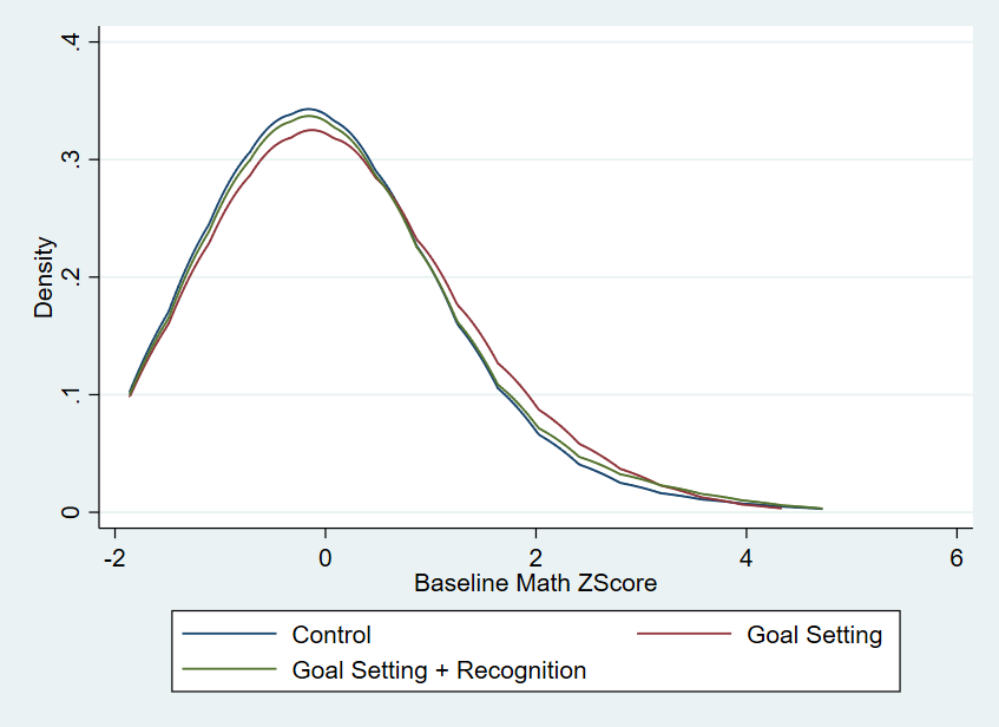
Table 14: Main Results - By Gap Between Expected and Actual Baseline Score (Degree of *Overestimation*)

	(1)	(2)	(3)	(4)	(5)	(6)	(7)
Dependent Variable:	Time-use Index	Effort Index	Self-discipline Index	Confidence Index	Aspiration Index	Math Z-score	English Z-score
	Panel A: Overestimating Baseline Performance (Gap $\geq$ points)						
Goal-Setting	0.096**	0.116**	0.103**	0.018	-0.008	0.092	0.010
	(0.046)	(0.052)	(0.046)	(0.035)	(0.040)	(0.069)	(0.067)
Goal-Setting + Recognition	0.074	0.087	0.079*	0.022	-0.023	0.041	0.024
	(0.047)	(0.057)	(0.044)	(0.043)	(0.045)	(0.068)	(0.064)
	(0.055)	(0.067)	(0.052)	(0.050)	(0.052)	(0.096)	(0.067)
Observations	9,908	9,254	10,186	10,114	9,421	10,480	10,480
Baseline Outcome Controlled	Yes	No	No	No	Yes	Yes	Yes
	Panel B: Not Overestimating Baseline Performance (Gap $<$ 10 points)						
Goal-Setting	0.133***	0.074	0.061	-0.003	0.065	0.033	0.156**
	(0.047)	(0.051)	(0.044)	(0.039)	(0.043)	(0.100)	(0.074)
Goal-Setting + Recognition	0.145**	0.014	0.068	-0.053	-0.021	0.040	0.127*
	(0.058)	(0.057)	(0.045)	(0.045)	(0.051)	(0.094)	(0.071)
Observations	8,621	8,111	8,830	8,805	8,276	9,085	9,085
Baseline Outcome Controlled	Yes	No	No	No	Yes	Yes	Yes

Notes: This table reports the heterogeneity (by level of overestimation of baseline performance) of the impact of interventions on main outcomes of interest: Time-use Index, Effort Index, Self-Discipline Index, Confidence Index, Aspiration Index, Math test score, and English test score. Construction of these indices is discussed in Section 2. Panel A has the sample of only male students while Panel B reports the results for the sample of female students. Standard errors are clustered at the level of school. *p<0.1; **p<0.05; ***p<0.01.

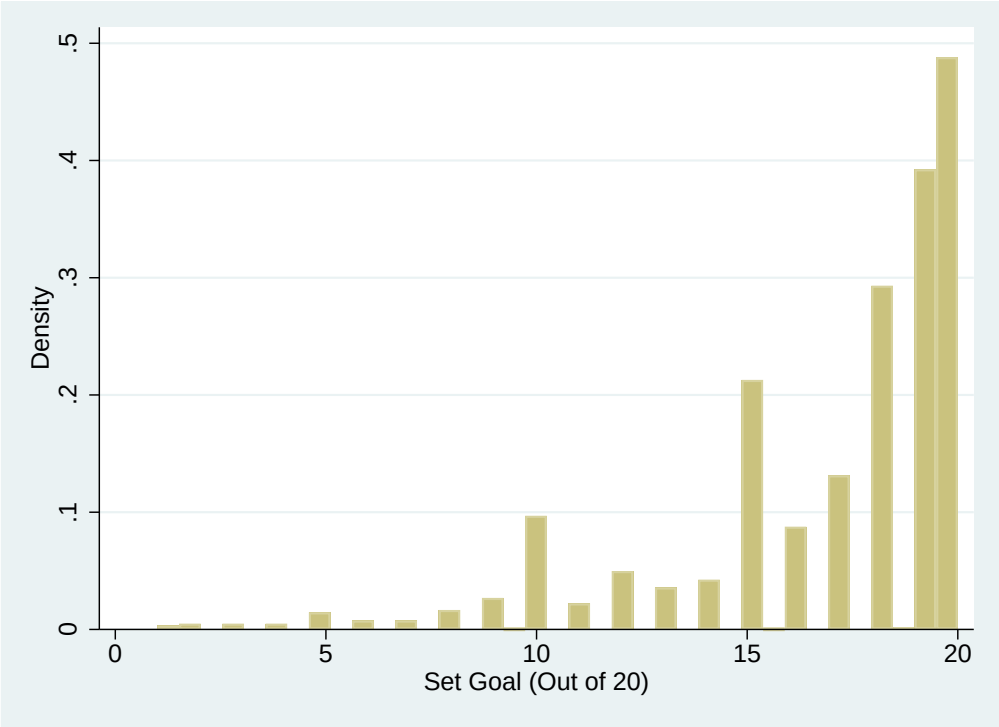
Figures

Figure 1: Distribution of Baseline Math Score



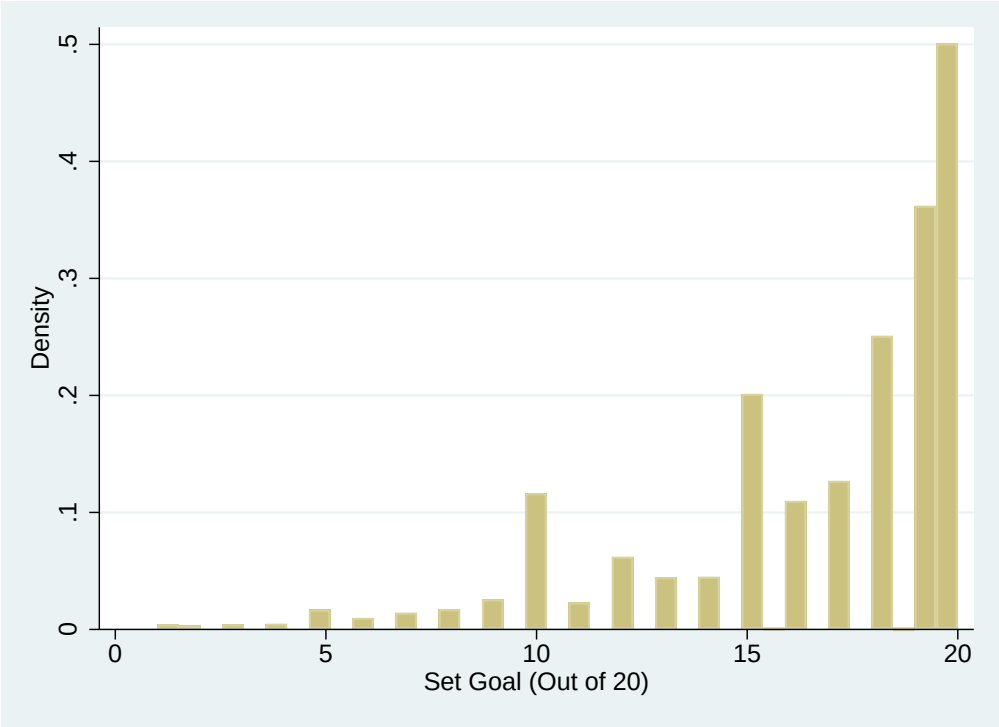
Notes: This figure shows the kernel density of the Math Z-score from the test conducted at the Baseline. As shown by the balance checks in the Table 3, these scores are statistically similar to each other in comparison across the three study groups.

Figure 2: Distribution of Goals: Goal-Setting Arm



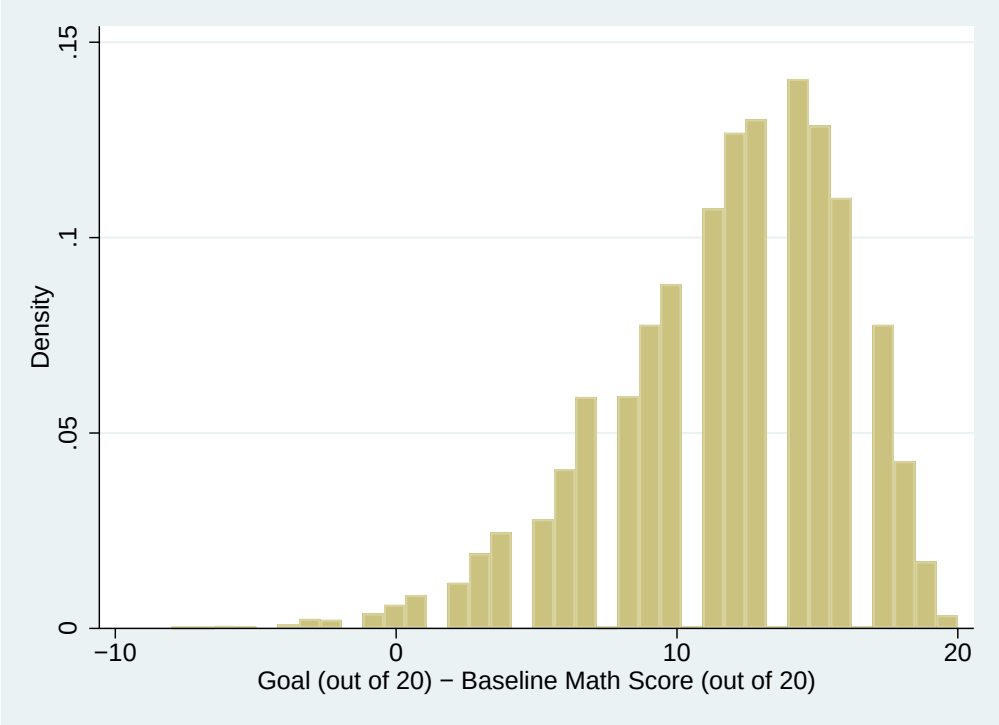
Notes: This figure shows the distribution of set goals (out of 20) for all the students in the Goal-Setting only treatment arm.

Figure 3: Distribution of Goals: Goal-Setting + Recognition



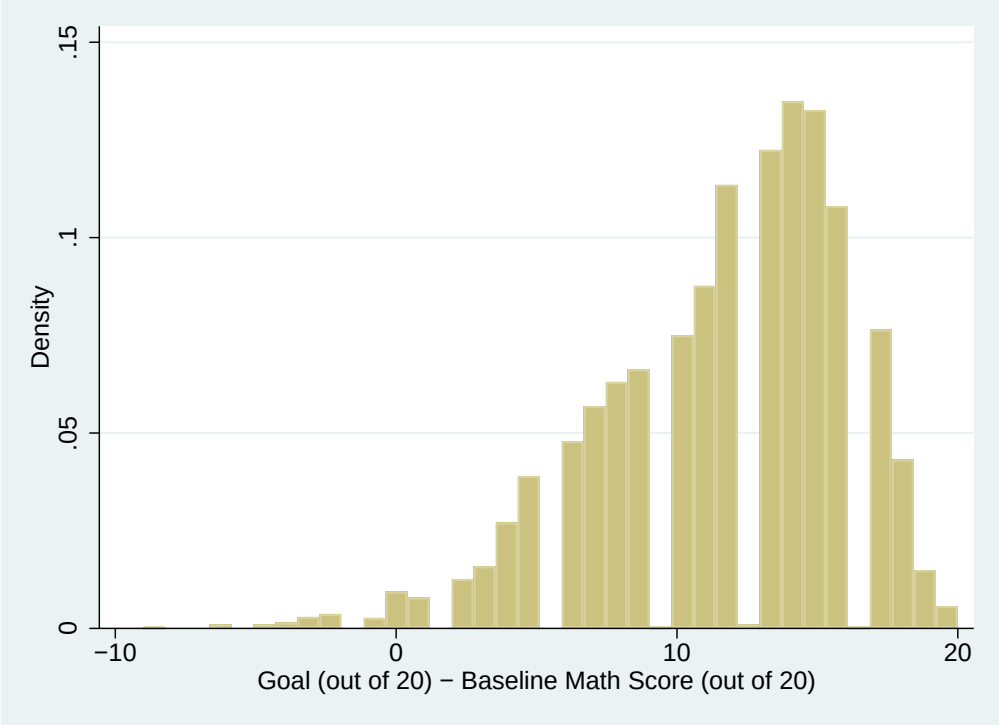
Notes: This figure shows the distribution of set goals (out of 20) for all the students in the Goal-Setting + Recognition treatment arm.

Figure 4: Goal *minus* Actual Baseline Score: Goal-Setting Arm



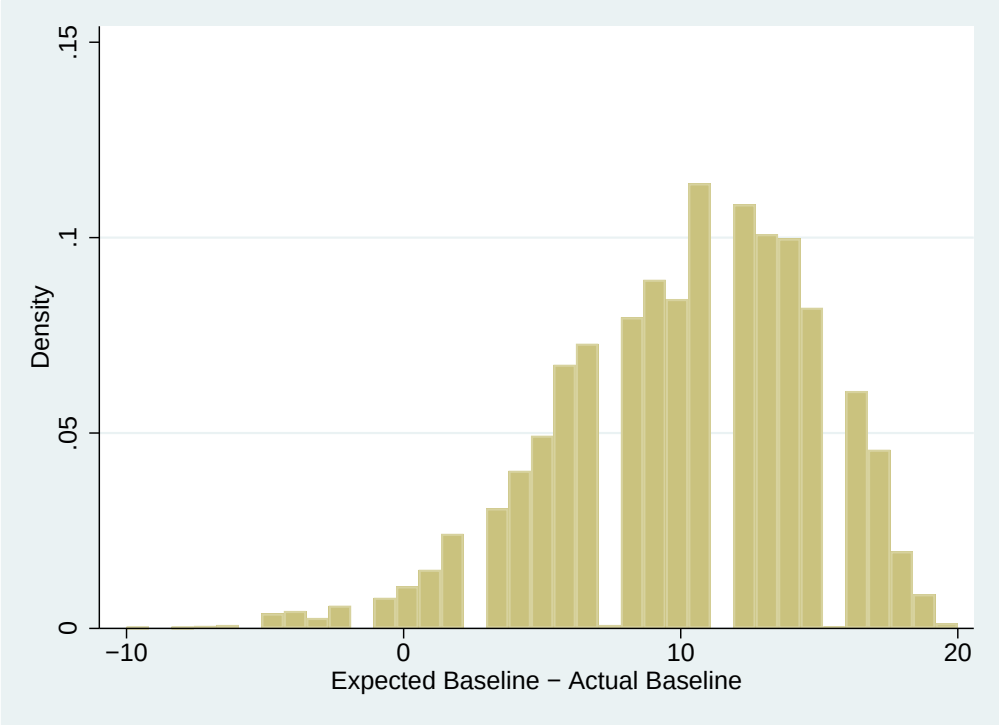
Notes: This figure shows the distribution of the difference between the set goal and actual baseline score for all the students in the Goal-Setting only treatment arm.

Figure 5: Goal *minus* Actual Baseline Score: Goal-Setting + Recognition



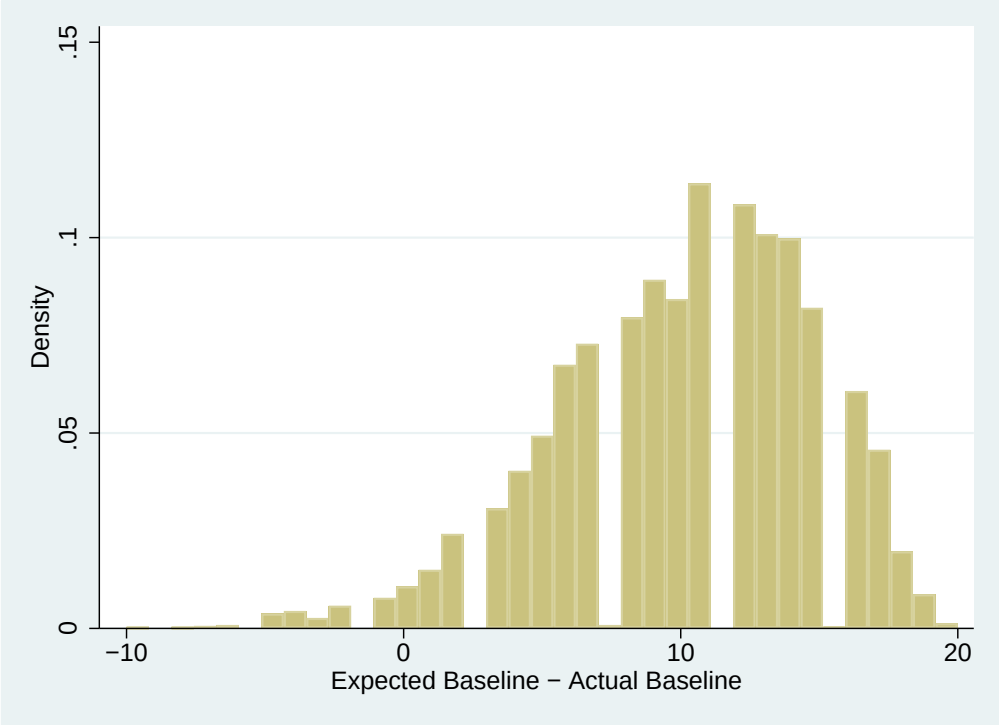
Notes: This figure shows the distribution of the difference between the set goal and actual baseline score for all the students in the Goal-Setting + Recognition treatment arm.

Figure 6: Expected *minus* Actual Baseline Score: Goal-Setting Arm



Notes: This figure shows the distribution of the difference between the actual and expected baseline score for all the students in the Goal-Setting only treatment arm.

Figure 7: Expected *minus* Actual Baseline Score: Goal-Setting + Recognition



Notes: This figure shows the distribution of the difference between the actual and expected baseline score for all the students in the Goal-Settings + recognition treatment arm.

Appendix: Tables

Table A-1: Overestimating and Overambitious Students

	Overambitious	
Overestimating	Yes	No
Yes	0%	60%
No	5.8%	34%

Notes: This table reports the cross tabulation of the fraction of treated students (in both treatment arms) who are *overestimating* their baseline performance and/or set overambitious goals. The definition used to construct overestimation and overambitious cutoff is discussed in section 2.5.

Table A-2: Main Results (Pooled Treatments)

Dependent Variable:	(1) Time-use Index	(2) Effort Index	(3) Self-Discipline Index	(4) Confidence Index	(5) Aspiration Index	(6) Math Z-score	(7) English Z-score
Treatments Pooled	0.107***	0.080*	0.080**	0.005	-0.003	0.042	0.063
	(0.037)	(0.044)	(0.037)	(0.032)	(0.034)	(0.062)	(0.053)
Observations	12,715	11,908	13,049	12,981	12,145	13,426	13,426
Baseline Outcome Controlled	Yes	No	No	No	Yes	Yes	Yes

Notes: This table reports the impact of pooled interventions (both Goal-Setting and Goal-Setting + Recognition groups pooled together) on main outcomes of interest: Time-use Index, Effort Index, Self-Discipline Index, Confidence Index, Aspiration Index, Math test score, and English test score. Construction of these indices is discussed in Section 2. Standard errors are clustered at the level of school. *p<0.1; **p<0.05; ***p<0.01.

Table A-3: Parent's and Teacher's Efforts

Dependent Variable:	(1) Parent's Effort Index	(2) Teacher's Effort Index
Goal-Setting	0.004	0.038
	(0.048)	(0.078)
Goal-Setting + Recognition	-0.015	0.071
	(0.046)	(0.090)
Observations	13,183	13,113
Baseline Outcome Controlled	No	No

Notes: This table reports the impact of interventions on Parents Effort Index and Teacher's Effort Index. Construction of these indices is discussed in Section 2. Standard errors are clustered at the level of school. *p<0.1; **p<0.05; ***p<0.01.

Table A-4: Minimum Detectable Effect Sizes

Outcome	(1) Observed Effect Size	(2) Minimum Detectable Effect Size
Time-Use Index	0.113	0.086
	(0.042)	
Effort Index	0.106	0.093
	(0.048)	
Discipline Index	0.09	0.079
	(0.043)	
Confidence Index	0.018	0.098
	(0.034)	
Aspiration Index	0.018	0.102
	(0.039)	
Math Z score	0.056	0.181
	(0.071)	
English Z score	0.062	0.193
	(0.065)	

Notes: This table reports the observed effect sizes (as a fraction of standard deviation) along with the standard errors and the retrospectively calculated minimum detectable effect size using the endline distribution of the outcome variables of the control group. The chosen parameters are $\alpha = 0.05$, power $= 0.8$. Value of intra cluster correlation ρ is taken from the control group distribution of each outcome at endline. *p<0.1; **p<0.05; ***p<0.01.

Table A-5: Attrition Rate

Study Group	(1) Attrition at Endline
Control	28.25%
Goal-Setting	26.72%
Goal-Setting + Recognition	24.07%

Notes: This table reports the attrition rates observed at endline survey for each of the study groups. *p<0.1; **p<0.05; ***p<0.01.

Table A-6: Attrition and Baseline Characteristics

	(1)	(2)
Baseline Variable	Coefficient	P-Value
Gender (Girl = 1)	-0.157	0
	[.017]	
Living with parents = 1	-0.001	0.931
	[.009]	
Mother's Occupation is Farming	-0.024	0.123
	[.015]	
Mother is housewife	-0.001	0.882
	[.008]	
Mother's occupation (Other non farming)	0.003	0.78
	[.011]	
Father's occupation is Farming	-0.021	0.1
	[.012]	
Father has no occupation	0.004	0.11
	[.003]	
Father's occupation (non farming)	-0.006	0.675
	[.014]	
Mother can read and write in English = 1	-0.011	0.379
	[.012]	
Father can read and write in English = 1	0.011	0.336
	[.012]	
Number of people in household	-0.005	0.683
	[.011]	
Asset Index	0.08	0.001
	[.024]	

Notes: This table reports the predictors of attrition using baseline characteristics of students. The dependent variables are baseline characteristics and the independent variable is a dummy taking the value 1 if student attrited at endline and 0 otherwise. Standard errors are clustered at the level of school. *p<0.1; **p<0.05; ***p<0.01.

Table A-6: Attrition and Baseline Characteristics *Continued*

Baseline Variable	(1) Coefficient	(2) P-Value
Student repeating current grade = 1	-0.015	0.002
	[.005]	
Student new to school (if not repeating grade) = 1	-0.018	0.02
	[.007]	
Student remembers last year's Math score = 1	0.002	0.896
	[.015]	
Student attended special session for last math exam = 1	0.01	0.448
	[.014]	
Student studied and did homework outside school	-0.127	0
	[.024]	
Helped in household	-0.029	0.222
	[.024]	
Sleeping frequency	0.04	0.126
	[.026]	
Played games/spend time with friends outside school	0.159	0
	[.026]	
Time spent studying math outside school	-0.08	0
	[.022]	
Wants to pursue further education after graduating school	-0.043	0
	[.009]	
Math score at Baseline	-0.551	0
	[.097]	
Expected math score at Baseline	-0.261	0.037
	[.123]	

Notes: This table reports the predictors of attrition using baseline characteristics of students. The dependent variables are baseline characteristics and the independent variable is a dummy taking the value 1 if student attrited at endline and 0 otherwise. Standard errors are clustered at the level of school. *p<0.1; **p<0.05; ***p<0.01.

Table A-7: Lee Bounds on Treatment Effect

	1st Imputation			2nd Imputation			3rd Imputation			4th Imputation		
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)	(12)
Dependent Variable:	Time-use	Effort	Self-Discipline	Time-use	Effort	Self-Discipline	Time-use	Effort	Self-Discipline	Time-use	Effort	Self-Discipline
Goal-Setting	0.068**	0.055*	0.054*	0.059*	0.049	0.050	0.059*	0.042	0.045	0.054	0.036	0.041
	(0.033)	(0.030)	(0.030)	(0.033)	(0.033)	(0.032)	(0.033)	(0.033)	(0.031)	(0.033)	(0.033)	(0.030)
Goal-Setting + Recognition	0.059*	0.026	0.042	0.051	0.023	0.038	0.052	0.020	0.035	0.046	0.017	0.031
	(0.034)	(0.036)	(0.032)	(0.034)	(0.038)	(0.034)	(0.035)	(0.036)	(0.033)	(0.036)	(0.037)	(0.030)
Observations	17,337	18,281	18,281	17,337	18,281	18,281	17,337	18,281	18,281	17,337	18,281	18,281
Baseline Outcome Controlled	Yes	NA	NA	Yes	NA	NA	Yes	NA	NA	Yes	NA	NA
Shift Factor	0.4	0.4	0.4	0.45	0.45	0.45	0.5	0.5	0.5	0.55	0.55	0.55

Notes: This table reports the results of estimating lee bounds on the set of main outcomes which showed significant movements i.e. Time-use Index, Effort Index, and Self-Discipline Index. Section 4.2 explains the analytical process in detail. While the imputation of attrits is done starting at 0.05 shift factor for each group, due to space constraints, this table reports only four such imputations ranging from 0.4 to 0.55. *p<.1; **p<.05; ***p<.01.

A Appendix: Figures

Figure A.1: Script for "Goal Setting" Schools - Part 1 of 2

1.	**Check to confirm if it is a Treatment 1 school. VERY IMPORTANT**
2.	Give a summary of the exercise to the class teacher and ask for their assistance in carrying it out with the Form 2 students. You can say: *"We are doing a short exercise about student motivations and discussing personal goals with them'*
3.	Gather all students into the classroom and ask them to take a seat. Thank the students for their time after the test and the survey. Establish a relaxed environment where students are not afraid of speaking up and asking questions. Address the entire class about the importance of setting goals and achieving them. You can use the following script: *"Thank you so much for your time and patience. Hope you enjoyed your sodas. We would like to talk to you about the importance of setting goals and targets and striving towards achieving these targets. Setting goals is a fundamental component to long-term success. Goals help you focus and allocate your time and resources efficiently, and they can keep you motivated when you feel like giving up*
4.	Introduce the concept of Personal Best targets to students. You can say *"When setting targets or goals for yourself, it is important to make sure that the target is your personal best. For example, if you scored 85 out of 100 in your test, that means that you are capable of performing that well. Now when you set a goal for yourself for next year, should it be more than 85 or less? Yes, it should be more. Because you have once achieved 85 out of 100, it means you should push yourself to score higher. This higher score will be your Personal Best, meaning it is something you have yet to achieve."*
1.	Ask students to stand up from their seat. Tell the students to stand up and reach for the ceiling as high as they can. As they are reaching, say, *"Now reach three inches higher."* As students push themselves to reach higher, say, *"I thought you were already reaching as high as you could. Where did you get the extra three inches?"* Have the students sit down and respond to you about the following question. *"What is the difference between doing a good job and doing your personal best?"* Take answers from 1 or 2 students. If no one answers, say: *"When you do your personal best, it is even better than good. It is the best that you are capable of doing"*
2.	Now hand the questionnaires to the students and ask them to fill out their information. Make sure everyone knows how to do this.
3.	Turn to the first page of the questionnaire. You can say: *"We now want you to think about what you want to be when you grow up. Write this down on the sheet we provided you. [After 2 minutes]"* Make sure no one has any questions. *"Now we want you to think about how the knowledge of Math can help you become better at this profession. You can write as many things as you can think of, and they may be very small reasons. For example a carpenter needs to be good at Math to take accurate measurements"* Ask the teacher to go around the class and see if anyone is struggling to answer the question.
4.	Moving on to the goal setting questions: *"We now want you to think about how well you think you performed on the Math test you just took. The Math test had 20 questions so it was out of 20 marks. Think about how many questions you think you got right, and that should be your final score. Be as honest as possible. These scores will not be shared with your teachers or peers. Now we would like you to think about the exercise we did in trying to touch the ceiling, and how you pushed yourself to reach even higher after you set a higher goal for yourself. Think about how you study for your Math class and tests, and come up with ways of improving yourself to achieve an even higher target for yourself. You can do this by first looking at how well you think you performed on the Math test that you just took, and*

Figure A.1: Script for "Goal Setting" Schools - Part 2 of 2

	set a personal best (PB) target for next year's Math test which will be very similar to this test. Your personal best score should also be out of 20".
5.	Once the students have written down their Personal Best target/goals in Math, ask the class teacher to go over each student and check if the targets are reasonable. If not, the teachers should spend time with the student and go over their strengths to come to a more reasonable goal to achieve for them.
6.	For the Math final question, say: *"Now in the next question, write down what you scored in your Form 1 Math final school exam. The exam that you took was out of 100 marks. If you don't remember exactly what you scored, give an estimate, and we will check with your teachers"*
7.	Now ask students to fill out the rest of the questionnaire and let you know if they have any questions.
8.	Collect the students goal setting questionnaires. Before you finish say: "It is important that you remember your PB target score (out of 20) for any upcoming mathematics exams. Now that you have set this PB target, we encourage you to remember this target as you prepare for your exams, since these tests will help you achieve your target score in the next math test. During the school year, we will periodically go over the goals you set today, and see where you stand in your challenge to achieve them by the end of the year."
9.	Thank the students and excuse them.
10.	Ask teachers to put up the poster in the classroom or school as a reminder to the students. Encourage teacher to give reminders to students about their goal setting every month.

Figure A.2: Script for "Goal Setting + Recognition" Schools

1.	**Check to confirm if it is a Treatment 2 school. VERY IMPORTANT**
2.	Give a summary of the exercise to the class teacher, and ask for their assistance in carrying it out with the Form 2 students. You can say: "We are encouraging students to set higher goals for themselves. We want them to think about their score in last year's Math exam and set a personal best target for next year that is higher than that score. We will then track their performance to observe if it actually does improve. At the end of the year, those that meet their goals will be awarded at a grand ceremony at school."
3.	Carry out steps 2-8 same as that for Treatment 1 school
4.	Collect the students goal setting questionnaires. Before you finish say: "It is important that you remember your PB target score for any upcoming mathematics exams. Now that you have set this PB target, we encourage you to remember this target as you prepare for your exams, since these tests will help you achieve your target score in the math final exam. During the school year, we will periodically go over the goals you set today, and see where you stand in your challenge to achieve them by the end of the year. **At the end of the school year, students who meet their target will receive a certificate of recognition to be rewarded at a grand ceremony in the presence of their teachers, school peers, and some role models. Those that exceed their goals will receive a medal of "outstanding performance" at the ceremony, and those that don't meet their target will not receive any awards."**
5.	Thank the students and excuse them.
6.	Ask teachers to put up the poster in the classroom or school as a reminder to the students. Encourage teacher to give reminders to students about their goal setting every month.