All Pain and No Gain: When Goal Setting Leads to More Effort but No Gains in Test Scores *

Asad Islam[†] Sungoh Kwon[‡] Eema Masood[§] Nishith Prakash[¶]
Shwetlena Sabarwal[‖] Deepak Saraswat[**]

August 31, 2024

Abstract

We conducted an at-scale randomized control trial among 18,281 secondary students in Tanzania to examine the effects of self-set academic goals on students' efforts and academic outcomes. We also test the impact of combining goal setting with non-financial rewards. We find that goal-setting has a significant positive effect on self-reported student time use, study effort, and self-discipline along with a positive but statistically insignificant impact on the performance on the test. We also find that combining goal setting with recognition awards for achieving the goals does not demonstrate any complementary effects. Heterogeneity analysis suggests that goal-setting has a higher impact on the test performance for the students in the middle of the distribution of baseline learning levels. We also find that the impact of the treatment does not vary significantly across students' gender, socioeconomic background and by the type of chosen goals.

Keywords: Goal-Setting, Recognition Rewards, Student efforts, and Zanzibar.
JEL Codes: D9, I20, I25, O15, O55.

*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 L. Ross, Delia Furtado, Frank Schilbach, Samer Al Samarrai, Inaam Ul Haq 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), 17th Midwest International Economic Development Conference, NEUDC 2020, KDIS-DIME Conference and KDIS-ADB-HKUST 2021 Asia Impact Evaluation 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.**

[†]Monash University. asadul.islam@monash.edu.
[‡]Korea Institute of Public Finance. sokwon@kipf.re.kr.
[§]World Bank. eema107@gmail.com.
[¶]Northeastern University, CESifo, IZA, HiCN, GLO, and CReAM. n.prakash@northeastern.edu.
[‖]World Bank. ssabarwal@worldbank.org.
[**]Meta Inc. saraswat.d@gmail.com.

1 Introduction

A large number of initiatives to improve educational outcomes in developing countries have taken place, many of them such as school construction (Duflo, 2001), hiring additional teachers (Banerjee *et al.*, 2007; Muralidharan & Sundararaman, 2013), giving students textbooks (Glewwe *et al.*, 2009), classroom computers (Banerjee *et al.*, 2007), information (Islam, 2019), relieving the distance constraint (Muralidharan & Prakash, 2017; Fiala *et al.*, 2022), etc. require significant resource mobilization.[1] While there is significant progress in understanding the impact of resource-intensive interventions, the educational reality in Sub-Saharan African countries posits a fundamentally different challenge, especially given resource-poor settings. In this paper, we conduct an at-scale randomized control trial (RCT) among all the secondary schools in Zanzibar (an autonomous region of Tanzania) to test the impact of goal-setting (an inexpensive yet promising intervention) on student effort and academic achievement.

Goal setting is an inexpensive tool that is scalable without mobilizing significant resources if it works effectively.[2] Setting personal goals can act as a self-imposed commitment device to motivate oneself, increase effort, persistence, discipline, and self-regulation [see (Church *et al.*, 2001; Wiese & Freund, 2011)]. Furthermore, goal-setting can enhance student's interest in the subject matter (Rowe *et al.*, 2017), increase sensitivity to performance outcomes and prompt self-monitoring of performance attainments (Cleary & Zimmerman, 2004), promote student's self-efficacy in learning (Bandura & Schunk, 1981), and help individuals pursue a level of challenge that optimally exceeds their present capacity (Locke & Latham, 2002).

Previous studies in the psychology literature, such as (Locke, 1968) and (Locke & Latham, 1990), argue that goals can act as powerful motivators that may affect both thought and

[1] Overall, the results from these studies suggest that programs which may increase one academic outcome such as getting children into the classroom to improve school participation may not improve their test scores. These programs are also very resource-intensive. Moreover, they impose substantial costs on the government in resource-poor developing countries, making them difficult to scale up nationwide in poor developing countries.

[2] In addition to budgetary implications, the countries in Sub-Saharan Africa lack adequate human capital resources such as qualified teachers and tutors (Bank, 2012).

action towards improving an outcome. This idea also dates back to prospect theory, which suggests that goals can serve as reference points (Kahneman & Tversky, 1979), with the psychological motive of loss aversion causing individuals to strive to reach their goals (Heath *et al.*, 1999). In the context of education, goal-setting could motivate students to increase effort to achieve those goals, thereby improving their academic performance (see (Church *et al.*, 2001; Locke & Latham, 1990; Wiese & Freund, 2011), for examples).

It is conceivable that similar actions in the context of education can improve academic performance. Previous studies show that goal-setting is positively associated with growth mindsets, achievement, engagement, and academic outcomes (Burns *et al.*, 2017; Martin & Liem, 2010; Martin & Elliot, 2016b,a). Moreover, specific and challenging goals lead to better performance since these goals reduce the ambiguity of what is to be achieved (Locke & Latham, 2002). On the other hand, performance goals have also been linked to decreased self-efficacy (Middleton & Midgley, 1997; Skaalvik, 1997), thereby making the impact of goal setting on performance in education *a priori* ambiguous.

This study examines the impact of a goal-setting intervention that aims to improve students' academic outcomes in almost all public schools (187 schools) in Zanzibar (an autonomous region of Tanzania). We test the impact of goal-setting using two treatment groups. In the first treatment group (64 schools), we encourage students to set their own goals (a target score) for improvements in math test scores. In the second treatment group (61 schools), we add performance-based non-financial recognition awards (medals, certificates, backpacks, etc.) for achieving the goals. Such non-financial recognition awards provide social recognition from teachers, peers, or society. The recognition acts as an extrinsic incentive to work harder towards their goal. Economic theory suggests that extrinsic incentives can increase effort and achieve better outcomes [see (Besley & Ghatak, 2008; Ashraf *et al.*, 2014)]. When faced with extrinsic incentives and external recognition due to their improved performance, students may increase study effort and become more focused on time use and discipline. A combination of an extrinsic (recognition) and intrinsic incentive (self-set goal)

may have higher gains than 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; Heyman & Ariely, 2004; Islam *et al.*, 2020)].

In this paper, we attempt to answer two important questions. First, do self-set goals provide sufficient impetus for improvement in academic outcomes? Second, can recognition awards tied directly to goal achievement further enhance these outcomes? Improving the motivation of young students to work harder and perform better has been challenging (Clark *et al.*, 2020). Therefore, from a policy standpoint, goal-setting offers a low-cost, scalable option with intrinsic merit beyond its instrumental value in promoting student effort and academic achievement.

We find that self-set goals lead to a significant positive impact on self-reported time use, student effort, and self-discipline, but it is hard to uncover if these movements signify a real improvement in outcomes or are driven by social desirability. However, we find improvement in the performance on the math test, although it is not statistically different from zero (0.04 SD with p-value = 0.5). These results are averaging over all students, but the students with different pre-existing ability levels may respond differently to a softer behavioral intervention. While our study is not equipped to measure these differences rigorously, we attempt to document any suggestive evidence hinting at these differences in our heterogeneity analysis. We find that students in the *middle* of the distribution of baseline (pre-intervention) math scores demonstrate relatively better performance on the endline math test than their higher and lower performing counterparts. Since different students are likely to respond differently to the intervention, we also explore heterogeneity by students' gender, socioeconomic background and by the types of goals (reasonable or unreasonable goals). We proxy the students' socioeconomic background by examining whether either of their parents can read/write in English. Results suggest no statistically significant differences by gender of the treated students and that the students with parents who cannot read and write English per-

form statistically similarly to those with parents who can read and write English. However, we do find that the students who set very high goals end up performing better in the endline test as compared to their peers who set goals that are closer to their baseline performance.

This paper contributes to several related works of literature. Our findings contribute to the literature using experiments to estimate the impacts of self-set 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 [see, for example, (Clark *et al.*, 2020; Lent, 2018; Morisano *et al.*, 2010; Levitt *et al.*, 2016; O'Neil *et al.*, 1995)]. Among the closest to our study, (Clark *et al.*, 2020) in the context of undergraduate students in a public university in the US finds that only the goals specific to certain academic tasks show improvements in completion and performance. On the other hand, in the 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.*, 2019), 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 challenged to set a higher goal performed significantly worse than those in the goal treatment.

A closer geographic comparison comes from (Mbiti *et al.*, 2019) who test the impact of resource-intensive interventions (grants and incentives for teachers) on student performance in Tanzania. The results we document on the test performance and specifically in the *middle* of baseline math ability are not as large as (Mbiti *et al.*, 2019) but not trivial in size either, especially when comparing a resource-intensive set of interventions to a softer behavioral

nudge. 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 the very high rate of student dropout around this age.[3]

We also make small contributions to the very few empirical papers that have analyzed the role of 'status' and 'social recognition' (Ball *et al.*, 2001; Markham *et al.*, 2002; Charness *et al.*, 2010; Kosfeld & Neckermann, 2010). Although in this paper we do not directly test the pure 'status' dimension of awards and student recognition as predicted by many theories, we estimate if such awards complement the impact of goal-setting on students' academic performance, especially if tied directly to goal achievement.[4]

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 roll-out, it does not happen as often due to governmental and bureaucratic constraints. An intervention as cost-effective as goal-setting is easier to roll out at a larger scale and is better tested at such a large scale. In addition, 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 scale versus at a larger scale (Muralidharan & Niehaus, 2017).

[3]In Zanzibar, almost half of the students entering secondary schools drop out before 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.

[4]There are a large number of theories on status awards and social recognition and predictions in economics (e.g., (Frey, 1994, 2007; Besley & Ghatak, 2008) shows how external intervention may influence volunteers' efforts and performance. (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.

2 Experimental Design

2.1 Context

Zanzibar, off the coast of mainland Tanzania, is a semi-autonomous archipelago that comprises 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 approximately two-thirds living in Unguja. The literacy rate, defined by the percentage of people above ten 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 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 grade 5 onwards; subsequently, 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 especially low. At the lower secondary level, only half of all students passed the Form 2 exam (lower secondary level or grades 8 and 9), while the rest comprised 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 45% in 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 Timeline and Data Collection

We conducted the nationwide experiment in Zanzibar, where all grade eight students in public secondary schools were a part of the study sample. There were 187 secondary schools randomly assigned to two treatments and one control group. Table 1 shows sample sizes for each group, and Table 2 shows the timeline of the study, interventions, and reminders.

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 in English and Math. A curriculum-based standardized assessment was developed for this study. At the end of the data collection, the enumerators were instructed to make announcements to the two treatment groups on goal-setting exercises. Then, students in the treatment groups were given a (iv) Treatment Sheet to record their goals for the forthcoming Math test approximately nine months later.

Follow-up visits were conducted in August 2016, two months prior to the endline assessments. During these visits, enumerators provided reminders about the goal-setting activity conducted in February and noted participants to the impending endline test scheduled for the conclusion of the academic year. Additionally, in incentive schools, participants were further incentivized with reminders emphasizing that the attainment of set goals would result in non-financial awards.

Endline data collection was conducted in mid to end of October 2016, which included: (i) Survey with the Head Teacher, (ii) survey with teachers, and (iii) student survey and Assessment in English and Math. Only students from the baseline were tested in the endline.

2.3 Intervention

The goal-setting exercise in both treatment arms is motivated by (Martin & Elliot, 2016a,b) and made contextually implementable.[5]

The *Treatment 1* group received the goal-setting intervention. In this group, the enumerators introduced the concept of setting goals and used an interactive exercise to ensure students understood the concept of setting a goal. This exercise is described in Appendix Figure A.1. Before setting their personal best goals, students in the treatment groups were guided to write down what they would like to be when they grow up, the skills they would need in this dream profession, and how their knowledge of Mathematics would help them perform better in this profession. Then, students were asked to predict their scores in the baseline Math test and set their personal best goals for a similar exam at the end of the year. After setting their targets, students received guidance on enhancing their performance in the Math test, which included activities such as taking comprehensive class notes, actively engaging in class discussions, diligently completing homework assignments, and devising a structured study schedule. Furthermore, the individual teachers were asked to go around the classroom and check each student's goal individually to ensure that self-set targets are: (i) non-zero and positive, (ii) exceed the score students think they scored in the baseline Math test they just took, and (iii) no more than the maximum score possible.

The *Treatment 2* group, also known as "Goal-setting + Recognition," received the goal-setting intervention as in Treatment 1, and a non-financial recognition reward was announced for students achieving their goals. These rewards were in the form of certificates of achievement given in a ceremony in front of the whole school. Students were aware of this reward as part of the given script in Treatment 2 schools (see Appendix Figure A.2).

Teachers and headteachers in the treatment groups were asked to give students periodic reminders of their goals. Schools also received a poster to display, reminding students about the goal-setting activity and upcoming test.

[5]The exercise is described in Appendix Figure A.1 and Appendix Figure A.2.

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. There are no statistically significant differences between the treatment and control students on the baseline test score. Out of 26 total comparisons, there are slight imbalances on only two occasions in student characteristics. In addition, we do not find any statistically significant difference in school-level characteristics.

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 seven 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 1 and 2 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. An OLS regression of the target goal (a score out of 20) on an indicator for each of the treatment arms also confirms

that the goals are statistically similar across both the arms.[6] In an attempt to understand the goal-setting in detail, we plot the distribution of the gap between the set goal and actual baseline score for both the treatment arms in Figures 3 and 4 respectively. As observed, most students have set very high goals compared 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.

2.6 Outcome Measures

We analyze the impact of the intervention on *six* of the following outcome measures. Except the test scores, all other outcomes are compiled using student self-reports, and are therefore, not immune from a social desirability bias. We take that into account while interpreting the results from the self-reported outcomes. We discuss the construction of these outcome measures below in detail:

Test Score: The goal-setting exercise in both the treatment arms was 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 endline and baseline scores.[7] We also report similar z-scores for the English test, which were administered during baseline and endline.

Student Time-Use: In both the baseline and endline survey, we collected data on time use on an average weekday in 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*,

[6] The coefficient measuring the average difference between the goals set for treatment 1 and treatment 2 is -0.22 points, with a $|t| = 0.75$. The magnitude of the absolute difference is 1% of the mean goal score and the difference itself is statistically not distinguishable from 0.

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

and *hours studying math outside school*. Responses to these questions in the survey are coded on an increasing scale of 1 to 5, with 1 being the lowest and 5 being the highest value.[8] Standardized values of responses to all these questions are converted to a single Anderson's Index [see Anderson (2008)], called *Time-use Index*.[9]

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 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*. Students ranked these statements 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 students' 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*. Students ranked these statements 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 students' responses to statements measuring their 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 questions in class*. Students ranked these statements 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*.

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

[9] 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.

Aspirations Index: We collected students' 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.*[10] Students ranked these statements 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*.

Parent's and Teacher's Efforts Index: In the endline survey, we ask students questions related to teacher and parent effort, and we combine them to form indices for teacher and parent effort.[11] Questions related to the teacher's effort are: *Did your teacher assign any homework last week?*, *Did your teacher give quizzes or tests last month?*, and *If you had questions or problems, could you discuss them with your teacher freely?* Questions related to parent's efforts are: *During the last week, have your parents asked about your school life?*, *During the previous 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*.

2.7 Estimating Equation

We are interested in estimating the impact of goal-setting (Treatment 1: GS) and the 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.

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

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

Y_{is}^{Pre} is the baseline value of outcome observed at the endline, whenever available. β_1, and β_2 are our main coefficients of interest that highlight 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 Predictors of the test scores

The set of outcomes such as time-use, effort, self-discipline, confidence, and aspirations 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). However, which of these *intermediary outcomes* are strongly correlated with test scores is an empirical question in our setting. To understand these outcomes' relative and absolute importance, we examine the correlation between these outcomes and the test scores using the endline data for only the control group. Since the control group was not exposed to the goal-setting exercise, and the testing and evaluation were done parallel to other treatment groups, the correlations between intermediary outcomes and test scores are a good proxy of underlying counterfactual correlation for the treated groups. We regress the math z-scores of the control group from the endline on each of these intermediary outcomes. We also include as controls, the baseline values of the time-use and aspirations index.[12] Table 5 summarizes the results. Since the performance on the test is also impacted by a student's demographic characteristics, we sequentially include the gender of the student, an index of household assets (from baseline), and the English language ability of the parents as controls. Column 4 shows that time use and aspirations are strongly correlated with performance on the math test. Self-discipline and effort also positively impacts the test scores but the coefficients on

[12]The baseline survey measured only time-use and aspirations.

these covariates are not statistically different from zero.

3.2 Average Treatment Effects

We first present the estimates of the impact of goal-setting and goal-setting combined with recognition on our intermediary outcomes: *Time-use Index*, *Effort Index*, *Self-Discipline Index*, *Confidence Index*, and *Aspirations Index* in columns 1-5 in Table 6.[13] We present the estimates of the impact of the two treatment arms in Table 6 and find that both treatments led to a significant change in students' time use behavior, effort, and self-discipline. This analysis looks 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 ten comparisons to the false discovery test as per the Benjamin-Hochberg procedure [see (Benjamini & Hochberg, 1995)] and find that all results which show statistically significant movements pass the B-H test.[14]. Lastly, while the coefficients on both Math and English scores are positive, they are statistically insignificant for both the treatment arms.

While the goal-setting & goal-setting + recognition arms show differential movement in outcomes, these differences are not statistically different across arms (as shown by an F-test of difference in coefficients across treatment arms). Therefore, we pool both the treatment arms, and Table 7 shows the results.[15] Column 1 shows the overall effect of goal-setting on the *Time-use Index* by pooling both the treatments. The estimates suggest an aggregate effect of 10.7% of an s.d with smaller standard errors. Column 2 of Table 7 shows a positive and statistically significant impact (of 0.08 SD) of goal-setting on *Effort Index*. Column 3 shows a similarly positive and statistically significant impact of the treatment on *Self-Discipline Index*. In column 4, we analyze if goal-setting affected students' personalities by

[13] Apart from the baseline test scores, the baseline survey only recorded the self-reported measures of time-use and aspirations. Therefore, the baseline controls for only the time-use and aspirations are included in the estimating equation.

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

[15] The baseline survey measured time-use and aspirations, but not effort, self-discipline, and confidence. So, baseline outcomes are not controlled in columns (2), (3), and (4).

looking at the impact of the two treatments on a measure of student confidence: *Confidence Index*. We find that the impact of the goal-setting treatment on *Confidence Index* is positive but very small and statistically insignificant. In column 5, we analyze if goal-setting affected students' aspirations. We find little impact on *Aspiration Index*. The coefficient is very small and statistically insignificant. Although aspirations are very strongly correlated with the test scores in our setting, the goal-setting intervention was not able to move it, a finding that is consistent with the literature in psychology.[16]

The goal-setting intervention positively impacts the key intermediary outcomes identified in Section 3.1 (Table 5), which are strong predictors of the test scores. However, in our setting, it is hard to determine how much of these outcomes are driven by a social desirability influence versus a real behavioral change. Therefore, we rely largely on the test scores as the main outcome to measure the success of goal-setting treatment. We show the impact of the pooled treatment arms on the z-scores of endline Math test scores in column 6 of Table 7. We find that combining both the treatments led to a positive but statistically insignificant gain in test scores. We also re-define the test scores by combining the standardized values of both the Math and English test and find that the average improvement to be 0.05 of a SD but that is statistically not different from zero.[17] Improving test scores has not been trivial in the education literature. It could mostly be improved in studies testing expensive interventions, which, unlike behavioral interventions, directly impact the cost of getting an education or classroom instruction.[18] Only a handful of behavioral interventions have shown a positive impact on test scores.[19] Our results are consistent with (Oreopoulos & Petronijevic, 2019) and (Dobriyoni *et al.*, 2019) who *do not* find the impact of the social psychology interventions on academic performance in their studies in Canada.

[16]Literature in psychology demonstrates that aspirations are shaped early in a child's life and tend to decline and 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.

[17]The estimated coefficient on the combined test score is 0.05 with a standard error of 0.048.

[18](Muralidharan *et al.*, 2019; Muralidharan & Sundararaman, 2011; Fiala *et al.*, 2022) are examples of few such studies.

[19]Few notable examples include (Bettinger & Baker, 2014; Alan *et al.*, 2019).

While the impacts on self-reported outcomes are not completely immune from demand effects and social desirability bias, our study has certain features that reduce the gravity of these concerns. First, the endline for this study data was collected *nine* months after the baseline. This longer time gap helps to mitigate the potential for demand effects, as participants are less likely to remember the specific expectations or hypotheses of the researchers [see (Zizzo, 2010), (De Quidt *et al.*, 2018)]. Second, when collecting the endline data, students were reminded that their response are anonymous. Also, the students filled in their response to the endline survey themselves with apropriate privacy. Both these measures have been shown in the literature to be associated with less social desiarability bias [see (Tourangeau & Yan, 2007), (Nederhof, 1985)]. Finally, if the treatment effects were primarily driven by experimental demand, we would expect to see statistically significant impacts on all self-reported outcomes, including the confidence and aspiration index [see (Orne, 2017), (Nichols & Maner, 2008)]. The absence of such consistent effects across all self-reported measures suggests that demand effects are unlikely to be the primary driver of our results.

Overall, our results suggest that while goal-setting results in an improvement in learning outcomes, the improvement is measured rather imprecisely and might have significant underlying heterogeneities.

3.3 Heterogeneities

In this section, we explore potential heterogeneity in the impact of goal setting by math ability measured at the baseline, by gender, and socioeconomic status of the households.

3.3.1 Treatment effect by the baseline learning distribution

Students with differences in pre-intervention *ability* may respond differently to the intervention. While this study is not designed to formally measure the underlying systematic heterogeneity by ability, averaging over all ability levels might mask some interesting im-

pacts. For example, students with very high ability likely do not need a nudge to perform better in the endline exam and may not demonstrate any impact. Similarly, it is conceivable that students starting from very low ability levels might show significant marginal gains from a nudge compared to their counterparts. On the other hand, it is also possible that students with very low levels of ability may not demonstrate any gains from the nudge because a smaller behavioral nudge like goal-setting may end up being *too softer* of an intervention to move their outcomes. We utilize the baseline math test scores distribution to explore these possibilities.

Figure 5 shows the impact of the goal-setting intervention (pooled treatment groups) on students divided into quartiles of the baseline math z-score. The first quartile represents the students scoring highest on the baseline math test, and the fourth quartile contains the lowest scoring students. Students in the first and the last quartile representing the highest and lowest performers do not demonstrate any change in their endline math scores. Students in the middle of the baseline ability distribution (second and third quartile) show a hint of improvement in endline scores. The third quartile, in particular, shows a large improvement of close to 0.1 SD (p-value = 0.039). Figure 6 shows the impact on key intermediary outcomes identified in Table 5. With small differences, time-use improves for students in all the ability groups. Except for the lowest performers, all other students also show improvements in effort and self-discipline. Not only the impact sizes here are measured statistically imprecisely, the directional impacts do not align well with the heterogeneity in the test scores seen in Figure 5. In addition, considering that these outcomes are all self-reported, we find the results on these intermediary outcomes to be inconclusive and urge the readers to interpret them with caution.

However, the post-goal setting differences in the intermediary inputs between the treatment and control groups may explain a part or whole of the gains in the test scores for the students in the middle of the baseline math distribution. To explore this, we conduct a mediation analysis that aims to record the sensitivity of the test score result to including

an intermediary input as an explanatory variable. However, since the intermediary input are also impacted by the treatment assignment (the main explanatory variable), a causal interpretation of this analysis is problematic. We only use this analysis to get a directional signal about the role of intermediary inputs in explaining the treatment effect on the test scores. In addition, to further refine this directional analysis, we run it separately for groups of students who are in the 2nd and 3rd quartile of the baseline distribution of math score.

We run a series of models that start by looking at the unmediated effect on the math test score and add an intermediary input in each further iteration. If the difference in intermediary inputs is a key driver of the impact on test scores, we should expect to see the coefficient on test score attenuate with each key intermediary input control. We present the findings in Figure 7. Panel A and B represent the sample of students in the 2nd and 3rd quartile of the baseline math score, and Panel C combines these two samples. Each of these panels demonstrate that the intermediary inputs move the coefficient on the effect size, but not far enough (both in magnitude and direction) to explain a significant portion of the treatment effect. This exercise suggests that the potential differences in attributes such as time-use and aspirations do not account for the gains from the goal setting.

3.3.2 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 find mixed results by gender.[20,21] We analyze the heterogeneity in the treatment effect on math scores in Figure 8 (left panel). We find that female students demonstrate higher average improvement in math scores (closely missing statistical significance with p-value = 0.153), but this is not statistically different from their male counterparts. Figure A.3 shows the gender-split of the

[20] (Dobriyoni *et al.*, 2019) evaluate interventions related to goal-setting in the context of college education but do not explore the effects differentially by the gender of students. (Huillery *et al.*, 2021) evaluated a program in France targeted at improving student performance by developing their motivation, effort, and self-discipline and found a slightly higher impact on test scores for female children.

[21] (Muralidharan *et al.*, 2019) show the substantial overall effect of a tutoring intervention on students' test scores in urban India but do not find any differential impact by the gender of students.

impact on intermediary inputs. Consistent with the heterogeneity by baseline math ability, we find that both male and female students report positive impact on time-use, effort and discipline, with slightly higher impacts for female students.

3.3.3 Socio-economic status

Students from different socioeconomic backgrounds might demonstrate varying levels of motivation when subjected to the goal-setting treatment. (Dobriyoni *et al.*, 2019) find some suggestive evidence that students with English as their mother tongue gained more from goal-setting in the context of college education.[22] (Muralidharan *et al.*, 2019) find no differential impact by socioeconomic status for an intervention that leads to a substantial change in test scores. We analyze this by looking at *parents being able to read and write English*.

The middle panel of Figure 8 shows the impact of goal-setting on the math scores of students, separated by their parents' English speaking/writing ability. Students whose parents can either speak or write in English are categorized in the group "Yes". Students whose parents have higher language ability demonstrate some improvement in their math scores, which is slightly higher than their counterparts. However, neither the individual coefficients nor the difference is statistically different from zero. Similar to the gender-split, the middle panel of Figure A.3 shows that both group of students report a higher impact on time-use, effort and discipline.

3.3.4 Overambitious goals

How realistic or unrealistic the set goals are may determine how the goal-setting treatment impacts achievement. To explore this, we divide the treated students into two groups - "overambitious" and "all other goals". Overambition here is a function of the distance between the goal for the endline test and the baseline math score. Specifically, we take the distribution of the difference between the set goal and the baseline math score and calculate the goal as

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

a fraction of the baseline score. Students for whom the fraction is more than the median are classified as setting "overambitious" goals.[23] The median in this case is 2.6 times the baseline score.

The right-most panel of Figure 8 shows the achievement gains for the students who have set overambitious goals versus all other students. Since goal setting was conducted only for the treatment groups, the full control group is used as the comparison group in each sub-sample. We find that the math performance for students setting overambitious goals increases by 0.08 of a SD, while that of all other students is statistically indistinguishable from zero. Sub-sample level estimates are slightly imprecise (due to the split), but a qualitative difference in the direction of achievement gains is not trivial. This finding seems consistent with the literature in that over-ambition or overestimating one's own ability can be linked to higher risk-taking ability, which may result in better innovation (Hershleifier *et al.*, 2012). On the contrary, looking at the impact on intermediary inputs (right-most panel of Figure A.3), we see slightly higher self-reported time-use, effort and discipline, and a much higher improvement on self-reported aspiration by students setting less ambitious goals. While it is likely that students who set an ambitious goal underestimate their overall efforts given the size of their goal and the ground to cover to achieve that, this study is not equipped with measures that can help further disentangle that effect.

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 students' performance. In that sense, the treatment effect we observe on certain outcomes may result from teachers altering their

[23]For this analysis, we drop a small number of observations where the set goals are less than the actual baseline performance.

behavior in connection to the treatment and not from goal-setting per se. The same concern also holds for parents altering their input in their children's studies. Table A-1 (columns 1-2) estimates equation 1 with parents' and teachers' effort indices as outcomes and finds that both of them *do not* demonstrate a value statistically different from zero. Similarly, restricting the sample to only the students from the 2^{nd} and 3^{rd} quartiles of baseline math scores (columns 3-4) shows that parents' and teachers' efforts do not change in response to the treatment. This analysis attenuates the concern that the observed treatment effects result from altered parents' and teachers' efforts.

4.2 Attrition

As discussed in Section 3.1, we have attrition in the study from baseline to endline. Table A-2 shows that this attrition ranges from 24.07% in (Goals + Recognition) the treatment to 28.25% in the control group. However, balance checks in Table 3 show that attrition is not selectively different in treatment vs. control and across both treatments. Nevertheless, 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-3 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 than not attrits. Overall, it looks like the ones who did not participate in endline were worse in baseline academic indicators. A regression of an indicator of attritting (equal to one) on the baseline math score within each study group (control and pooled treatment) shows that attrits performed lower than non-attrits at the baseline (a coefficient of -0.20 with a p-value of 0.005 for control group and -0.21 with a p-value of 0.000 for the pooled treatment group). Overall, while we provide evidence against differential attrition across study groups, we also find evidence that those who attrit had

weaker baseline outcomes than non-attrits.

We calculate the bounds on treatment effects by using the bounding procedure suggested by (Lee, 2009). Table A-4 shows the upper and lower bounds on treatment effects for all outcomes, along with a 90% confidence interval for bounds. The lower bound occurs when the best-performing students in control schools attrit. We have shown that low-performing students drive attrition, and hence it is the upper bound which becomes more relevant in our case. Going by the upper bounds, we conclude that the attrition can *at best* lead us to underestimate these impacts.

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 that the test scores increase as a result of being exposed to goal-setting. However, these estimates are small and imprecise[24], such that neither zero effects nor much larger effects can be ruled out in the results in the aggregate sample. However, we find suggestive evidence that goal-setting has a heterogeneous impact on students with varying levels of ability, wherein students in the *middle* of the ability distribution demonstrate slightly better performance on the math test.

Recent experimental studies on *goal-settings* and *nudges* find it challenging to have an impact on test scores [see (Dobriyoni *et al.*, 2019; Oreopoulos *et al.*, 2018; Oreopoulos & Petronijevic, 2019)]. For example, (Oreopoulos *et al.*, 2018) finds 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. There are notable exceptions like (Huillery *et al.*, 2021) that find a positive impact of interventions targeted to improve students' growth mindset on test scores. An important difference is that (Oreopoulos *et al.*, 2018; Huillery

[24]In comparison, (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 significantly reduces inequalities in achievement if implemented early in students' academic careers.

et al., 2021) are conducted in a developed country. Incentives targeted to move students' academic performance in disadvantaged settings may have higher marginal gains than the developed country settings. This is likely because the students in former settings are starting from lower performance levels. (Mbiti *et al.*, 2019) provide a closer geographic comparison where they test the impact of resource-intensive interventions in the form of grants, incentives for teachers, and a combination of both in Tanzania. Our average impact size of 0.04 SD from the full sample (although statistically insignificant) lies between the impact sizes of 0.01 SD and 0.07 SD from the incentives alone and grants alone interventions of (Mbiti *et al.*, 2019). Although our study is not designed to formally measure a heterogeneous impact of goal-setting by the students' underlying ability, we do find some suggestive evidence that it is most likely to work for students who do not belong to the strongest or the weakest part of the ability distribution. Our study provides evidence that less resource-intensive interventions in the form of behavioral nudges like goal-setting also carry the potential to influence student performance in disadvantaged settings.

When we combine goal-setting with a recognition award in the second treatment, we find a weaker impact (although not statistically different from goal-setting only treatment) on outcome measures. This result is consistent with 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 social component attached to them.[25] 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.[26] Hence, it is plausible that such social comparisons might have diluted the goal-setting *bite* of the intervention.

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

[26] (Heyman & Ariely, 2004) shows that efforts in social markets are much less sensitive to compensation than in the 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 mis-perceived as a preference for rewards.

Overall, the results from this study suggest that goal-setting works to enhance the academic outcomes of students who do not belong to the weakest or the strongest part of the baseline learning distribution. Furthermore, 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, 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.

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.

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

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.

Banerjee, A.V., Cole, S., Duflo, E., & Linden, L. 2007. Remedying Education: Evidence from Two Randomized Experiments in India. *The Quarterly Journal of Economics*, **122(3)**, 1235–1264.

Bank, World. 2012. Tanzania Service Delivery Indicators. *Technical Report*.

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

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**.

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. 2020. Using Goals to Motivate College Students: Theory and Evidence from Field Experiments. *The Review of Economics and Statistics*, **102(4)**, 648–663.

Cleary, T. J., & Zimmerman, B. J. 2004. Self-regulation empowerment program: A school-based program to enhance self-regulated and self-motivated cycles of student learning. *Psychology in the Schools*, **41**(5), 537–550.

De Quidt, J., Haushofer, J., & Roth, C. 2018. Measuring and bounding experimenter demand. *American Economic Review*, **108**(11), 3266–3302.

Dobriyoni, C.R., Oreopoulos, P., & Petronijevic, U. 2019. Goal Setting, Academic Reminders, and College Success: A large-scale Field Experiment. *Journal of Research on Educational Effectiveness*, **12(1)**, 38–66.

Duflo, E. 2001. Schooling and Labor Market Consequences of School Construction in Indonesia: Evidence from an Unusual Policy Experiment. *American Economic Review*, **91(4)**, 795–813.

Fiala, Nathan, Garcia-Hernandez, Ana, Narula, Kritika, & Prakash, Nishith. 2022. Wheels of change: Transforming girls' lives with bicycles.

Frey, B.S. 1994. How intrinsic motivation is crowded out and in. *Rationality and Society*, **6(3)**, 334–352.

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

Glewwe, P., Kremer, M., & Moulin, S. 2009. Many Children Left Behind? Textbooks and Test Scores in Kenya. *American Economic Journal: Applied Economics*, **1(1)**, 112–135.

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

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.

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.

Huillery, E., Bouguen, A., Charpentier, A., Algan, Y., & Chevallier, C. 2021. The Role of Mindset in Education : A Large-Scale Field Experiment in Disadvantaged Schools. *SocArXiv*, **Web**(3).

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.

Islam, A. 2019. Parent-Teacher Meetings and Student Outcomes: Evidence from a Developing Country. *European Economic Review*, **111**, 273–304.

Islam, A., Malek, M.A., Tasneem, S., & Wang, L.C. 2020. Can public recognition award backfire? Field experimental evidence on the renention and performance of volunteers with social-image concerns. *Working Paper. Monash University.*

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.*

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.*

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.

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., & 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.

Mbiti, I., Muralidharan, K., Romero, M., Schipper, Y., Manda, C., & Rajani, R. 2019. Inputs, Incentives, and Complementarities in Education: Experimental Evidence from Tanzania. *Quarterly Journal of Economics*, **134**(3), 1627–1673.

Middleton, M., & Midgley, C. 1997. Avoiding the demonstration of the lack of ability: An under-explored aspect of goal theory. *Journal of Educational Psychology*, **89**, 710–718.

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

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., & Prakash, N. 2017. Cycling to School: Increasing Secondary School Enrollment for Girls in India. *American Economic Journal: Applied Economics*, **9**(3), 321–350.

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

Muralidharan, K., & Sundararaman, V. 2013. Contract Teachers: Experimental Evidence from India. *NBER Working Paper No. 19440*.

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

Nederhof, Anton J. 1985. Methods of coping with social desirability bias: A review. *European journal of social psychology*, **15**(3), 263–280.

Nichols, A. L., & Maner, J. K. 2008. The good-subject effect: Investigating participant demand characteristics. *The Journal of General Psychology*, **135**(2), 151–166.

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*.

Orne, M.T. 2017. On the social psychology of the psychological experiment: With particular reference to demand characteristics and their implications. *Sociological Methods*, 279–299.

Rowe, D. A., Mazzotti, V. L., Ingram, A., & Lee, S. 2017. Effects of goal-setting instruction on academic engagement for students at risk. *Career Development and Transition for Exceptional Individuals*, **40**(1), 25–35.

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.

Skaalvik, E.M. 1997. Self-enhancing and self-defeating ego orientation: Relations with task and avoidance orientation, achievement, self-perceptions, and anxiety. *Journal of Educational Psychology*, **89**(1), 71–81.

Tourangeau, Roger, & Yan, Ting. 2007. Sensitive questions in surveys. *Psychological bulletin*, **133**(5), 859.

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.

Zizzo, D.J. 2010. Experimenter demand effects in economic experiments. *Experimental Economics*, **13**(1), 75–98.

Tables

Table 1: Sample Size at Baseline

Study Group	(1) No. of Schools	(2) 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)				Difference & S.E.	
	(1)	(2)	(3)	(4)	(5)	(6)
	Overall	Control	GS	GS + R	GS vs. Control	GS + R vs. Control
Male Student	0.448	0.457	0.412	0.476	-0.045	0.019
	[.497]	[.498]	[.492]	[.499]	[.021]**	[.019]
Age	16.603	16.612	16.555	16.643	-0.057	0.031
	[1.284]	[1.27]	[1.321]	[1.26]	[.08]	[.075]
Is the student repeating the current grade?	0.064	0.071	0.053	0.068	-0.018	-0.003
	[.245]	[.256]	[.224]	[.252]	[.009]*	[.011]
Whether father can read and write in English	0.679	0.683	0.706	0.641	0.023	-0.042
	[.467]	[.465]	[.456]	[.48]	[.025]	[.03]
Whether mother can read and write in English	0.53	0.532	0.548	0.507	0.017	-0.025
	[.499]	[.499]	[.498]	[.5]	[.031]	[.036]
Asset Index	0	-0.007	0.057	-0.055	0.064	-0.048
	[.839]	[.843]	[.756]	[.916]	[.082]	[.086]
Household Asset Index	0	-0.007	0.065	-0.065	0.072	-0.058
	[.852]	[.866]	[.765]	[.919]	[.091]	[.095]
Baseline English Test Z-Score	0.043	0	0.121	0.011	0.121	0.011
	[1.038]	[1]	[1.064]	[1.053]	[.09]	[.096]
Baseline Math Test Z-Score	0.047	0	0.096	0.054	0.096	0.054
	[1.029]	[1]	[1.041]	[1.052]	[.108]	[.112]
Spend More Than 1 Hour in Math (Baseline)	0.377	0.372	0.388	0.371	0.016	-0.001
	[.485]	[.483]	[.487]	[.483]	[.017]	[.018]
Spend More Than 30 Minutes in Math (Baseline)	0.661	0.66	0.674	0.648	0.014	-0.011
	[.473]	[.474]	[.469]	[.478]	[.021]	[.022]
Baseline Effort in School (z-score)	0.006	0	0.059	-0.047	0.059	-0.047
	[.998]	[1]	[.948]	[1.047]	[.051]	[.054]
Absence in Endline Exam	0.266	0.282	0.267	0.241	-0.015	-0.042
	[.442]	[.45]	[.443]	[.428]	[.03]	[.03]
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 differences are reported by comparing GS (Goal-Setting), GS + R (Goal-Setting + Recognition) to the Control group. SEs and SDs are contained in square brackets. Standard errors are clustered at the level of school.

Table 4: Balance on School Characteristics

	Mean (SD)				Difference & S.E.	
	(1)	(2)	(3)	(4)	(5)	(6)
	Overall	Control	GS	GS + R	GS vs. Control	GS + R vs. Control
Total Students in F2	132.595	143.86	0.267	120.768	-10.896	-23.092
	[132.577]	[133.077]	[.443]	[95.06]	[25.145]	[25.031]
Total Qualified Teachers in F2	4.832	4.638	0.267	4.951	0.308	0.313
	[2.062]	[1.799]	[.443]	[1.76]	[.456]	[.443]
Student-Teacher Ratio in F2	28.228	29.151	0.267	27.341	-1.098	-1.811
	[18.734]	[17.943]	[.443]	[16.596]	[4.234]	[4.054]
Does this school have two shifts?	0.602	0.583	0.267	0.553	0.083	-0.031
	[.492]	[.5]	[.443]	[.504]	[.114]	[.115]
Form 2 Pass Rate in 2015 for English	50.99	50.108	0.267	51.351	1.474	1.243
	[26.525]	[26.85]	[.443]	[26.804]	[8.328]	[8.223]
Form 2 Pass Rate in 2015 for Math	44.153	38.796	0.267	47.194	8.678	8.399
	[26.707]	[28.251]	[.443]	[23.507]	[8.408]	[8.532]
Form 2 Pass Rate in 2015 for Science	48.731	46.033	0.267	53.597	1.67	7.564
	[25.522]	[25.47]	[.443]	[21.711]	[8.198]	[8.476]
Average Teaching Experience in Month	150.491	139.479	0.267	162.742	10.063	23.263
	[87.158]	[96.821]	[.443]	[87.522]	[16.339]	[16.699]
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 differences are reported by comparing GS (Goal-Setting), GS + R (Goal-Setting + Recognition) to the Control group. SEs and SDs are contained in square brackets.

Table 5: Predictors of the Endline Math Score: Using the Control Sample

Dependent Variable:	(1) Endline Math Z-Score	(2) Endline Math Z-Score	(3) Endline Math Z-Score	(4) Endline Math Z-Score
Time-use Index	0.108***	0.112***	0.111***	0.113***
	(0.028)	(0.029)	(0.027)	(0.027)
Effort Index	0.020	0.030	0.030	0.033
	(0.025)	(0.026)	(0.025)	(0.025)
Self-discipline Index	0.022	0.016	0.016	0.018
	(0.019)	(0.018)	(0.018)	(0.018)
Confidence Index	-0.056	-0.048	-0.048	-0.050
	(0.054)	(0.052)	(0.053)	(0.053)
Aspirations Index	0.117***	0.115***	0.114***	0.112***
	(0.034)	(0.034)	(0.033)	(0.033)
Constant	0.064	0.223	0.220	0.210
	(0.095)	(0.144)	(0.142)	(0.128)
Observations	3,892	3,892	3,892	3,853
Control for Baseline Values	Yes	Yes	Yes	Yes
Gender	No	Yes	Yes	Yes
BL Asset Index	No	No	Yes	Yes
Parent's Eng Ability	No	No	No	Yes

Notes: This table reports the results from an OLS regression of the endline math z-score on each of the intermediary outcomes: 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. The sample in this table is restricted to only the control group students participating in the endline exam. Controls for the baseline values of time-use and aspirations index are included in all models. Standard errors are clustered at the level of school. *p<0.1; **p<0.05; ***p<0.01.

Table 6: 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
BL Outcome Controlled	Yes	N/A	N/A	N/A	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)	[0.02]	-	-	-	-	-	-
Prob > F (Test of diff. by treatment)	(0.79)	(0.29)	(0.64)	(0.51)	(0.30)	(0.65)	(0.95)

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. Baseline outcomes are not controlled in columns (2), (3), and (4) because the level of effort, self-discipline, and confidence were not measured in the baseline survey. 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.

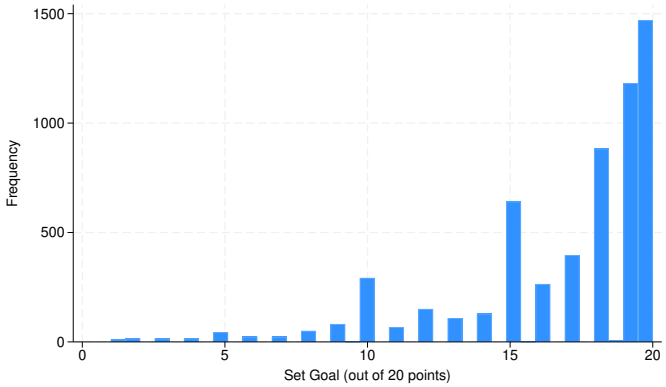
Table 7: Main Results - *Pooled Treatment Arms*

	(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 (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
BL Outcome Controlled	Yes	N/A	N/A	N/A	Yes	Yes	Yes

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. The main explanatory variable is the treatment indicator which pools both the GS only and GS + Recognition treatment arms. Construction of these indices is discussed in Section 2. Baseline outcomes are not controlled in columns (2), (3), and (4) because the level of effort, self-discipline, and confidence were not measured in the baseline survey. Standard errors are clustered at the level of school. *p<0.1; **p<0.05; ***p<0.01.

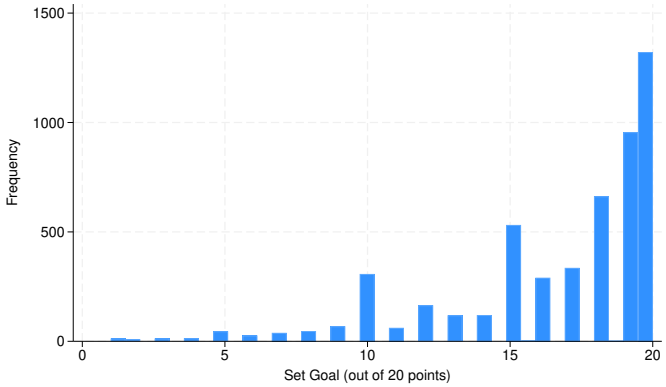
Figures

Figure 1: Distribution of Goals: Goal-Setting Arm



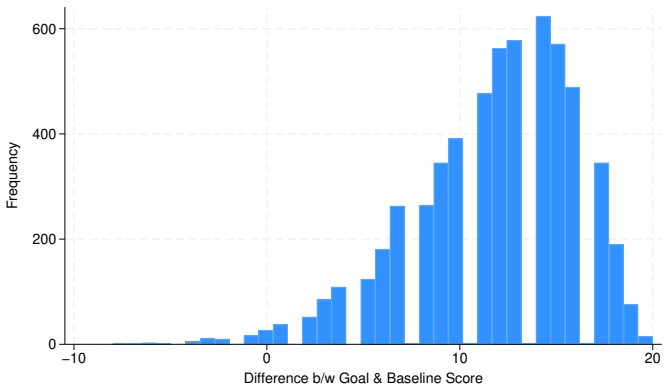
Notes: This figure shows the distribution of set goals (out of 20) for all the students in the Goal-Setting only treatment arm.

Figure 2: Distribution of Goals: Goal-Setting + Recognition



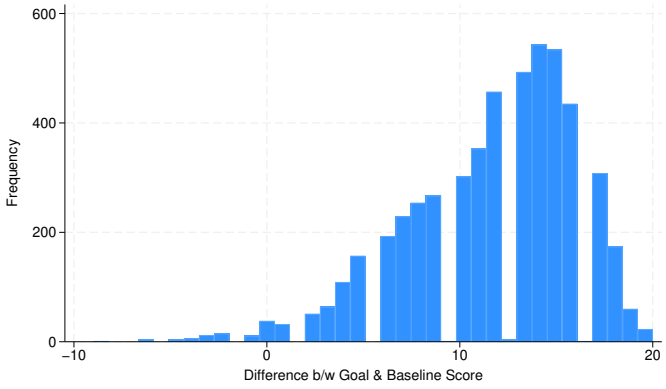
Notes: This figure shows the distribution of set goals (out of 20) for all the students in the Goal-Setting + Recognition treatment arm.

Figure 3: Goal *minus* Actual Baseline Score: Goal-Setting Arm



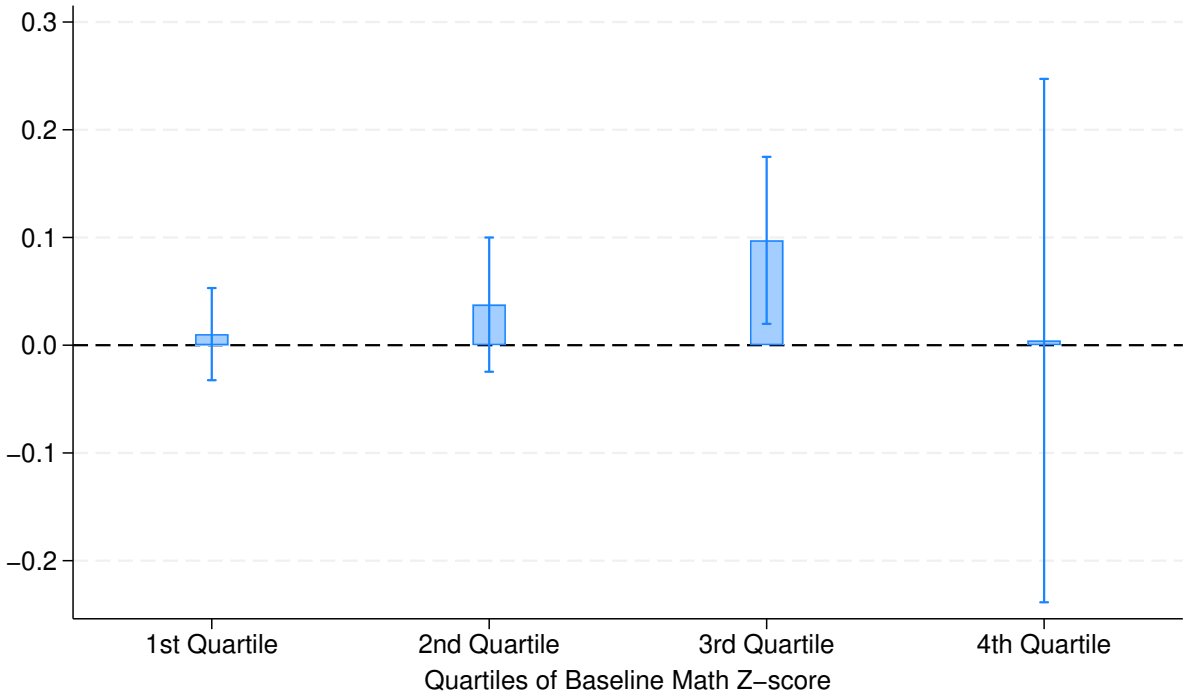
Notes: This figure shows the distribution of the difference between the set goal and actual baseline score for all the students in the Goal-Setting only treatment arm.

Figure 4: Goal *minus* Actual Baseline Score: Goal-Setting + Recognition



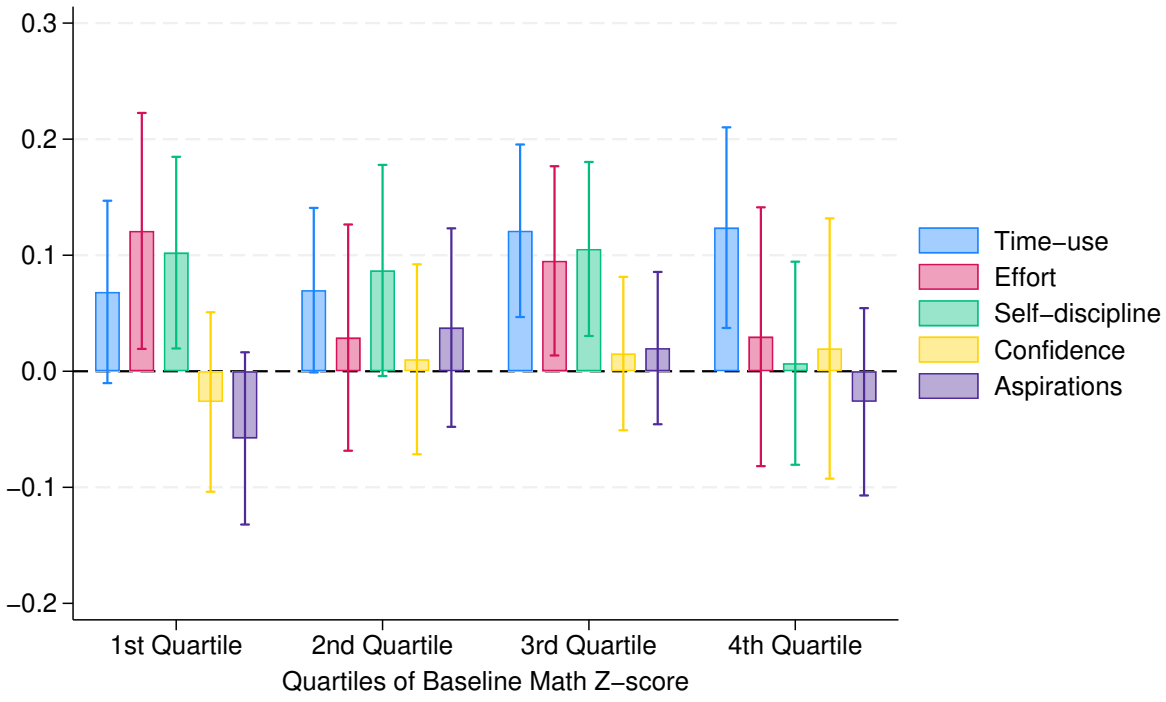
Notes: This figure shows the distribution of the difference between the set goal and actual baseline score for all the students in the Goal-Setting + Recognition treatment arm.

Figure 5: Heterogeneity in the Impact on Test Performance: By Baseline Math Score



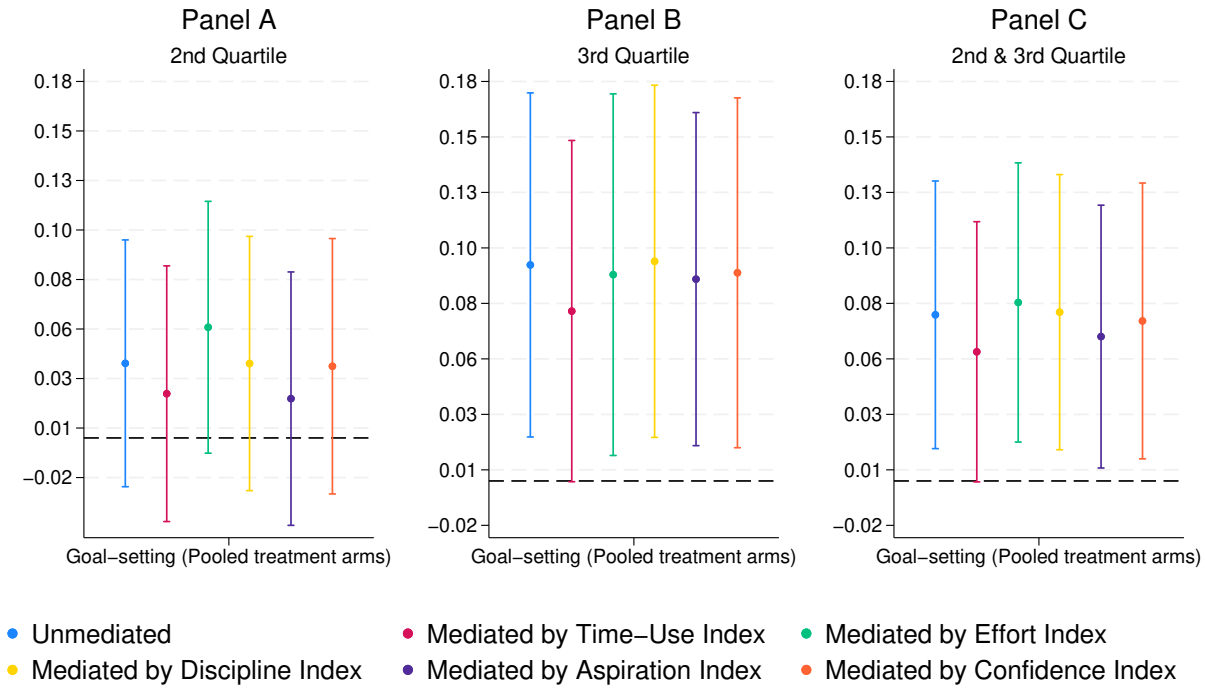
Notes: This figure shows the impact of goal-setting on the endline math test z-scores using equation (1). The sample is divided into quartiles of the distribution of baseline math z-scores. Standard errors are clustered at the level of school. Bars around the coefficient represent a 90% confidence interval.

Figure 6: Heterogeneity in the Impact on Intermediary Outcomes: By Baseline Math Score



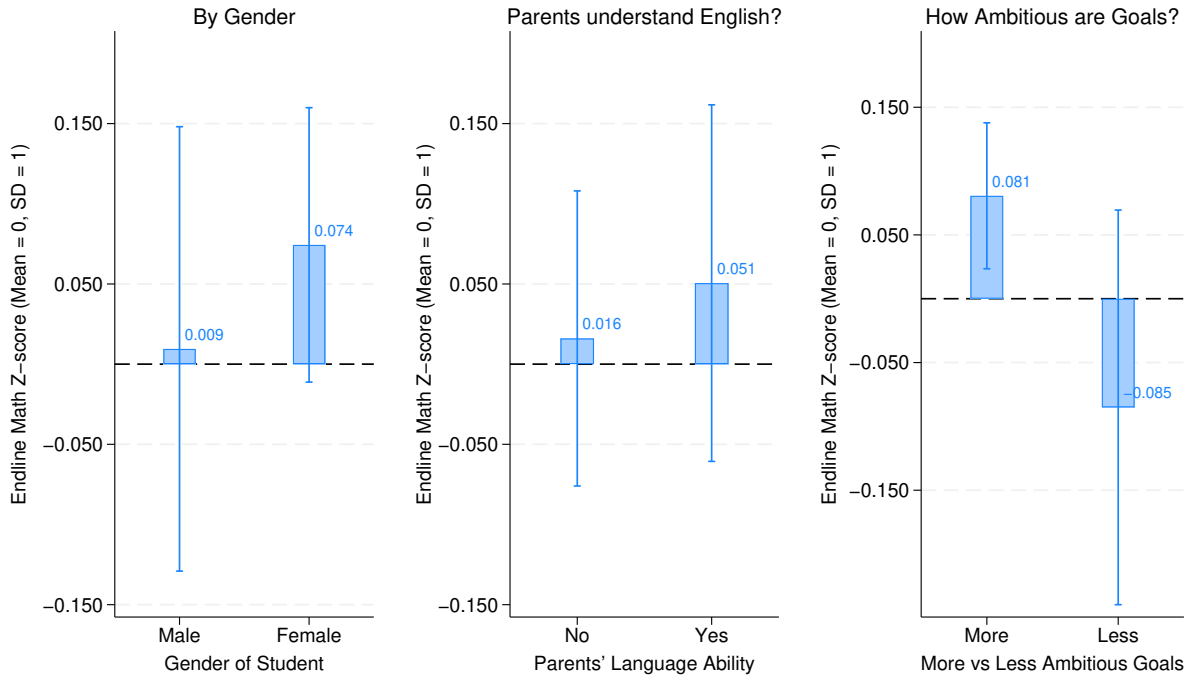
Notes: This figure shows the impact of goal-setting on intermediary outcomes (time-use, effort, self-discipline, and aspirations) using equation (1), where abseline values of all dependent variables are included as controls. The sample is divided into quartiles of the distribution of baseline math z-scores. Standard errors are clustered at the level of school. Bars around the coefficient represent a 90% confidence interval.

Figure 7: Mediation Analysis: For Treatment Effect in the Middle of the Baseline Math Ability



Notes: This figure summarizes the mediation analysis that explores if and how much of the treatment effect on math scores (for the middle distribution of baseline math scores) can be explained by intermediary inputs. Equation (1) is used to capture the treatment effect with further iterations where each of the intermediary inputs are used as controls. Standard errors are clustered at the level of school. Bars around the coefficient represent a 90% confidence interval.

Figure 8: Heterogeneity in the Impact on Test Performance: By Gender and Socioeconomic Background



Notes: This figure shows the impact of goal-setting on the endline math test z-scores using equation (1). The sample is divided by the gender (left panel), the english speaking/writing abolity of the students' parents (middle panel) and if their set goals are overambitious in nature (right panel). Goals that are 2.6 times or higher of the baseline math score are considered overambitious. Standard errors are clustered at the level of school. Bars around the coefficient represent a 90% confidence interval.

Appendix: Tables

Table A-1: Parent's and Teacher's Efforts

	(1)	(2)	(3)	(4)
	Full sample		*2nd & 3rd Quartile*	
Dependent Variable:	Parent's Effort Index	Teacher's Effort Index	Parent's Effort Index	Teacher's Effort Index
Goal-setting Pooled	-0.005	0.054	-0.005	0.065
	(0.040)	(0.073)	(0.047)	(0.078)
Observations	13,183	13,113	5,891	5,844
Baseline Outcome Controlled	No	No	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-2: Attrition Rate

	(1)
Study Group	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-3: Attrition and Baseline Characteristics - *(Continued)*

	(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-3: Attrition and Baseline Characteristics

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-4: Lee Bounds on Treatment Effect

	(1)	(2)	(3)	(4)
Index	**Treatment Effect**	**Lee Bound (Lower)**	**Lee Bound (Upper)**	**CI (90%)**
Time-Use	0.107	0.04	0.21	[0.01, 0.24]
	(0.037)	(0.024)	(0.026)	
Effort	0.08	0.08	0.165	[0.05, 0.20]
	(0.044)	(0.025)	(0.029)	
Discipline	0.08	0.032	0.162	[0.004, 0.21]
	(0.037)	(0.023)	(0.035)	
Confidence	0.005	0.005	0.111	[-0.02, 0.15]
	(0.032)	(0.022)	(0.029)	
Aspirations	-0.003	0.007	0.089	[-0.02, 0.13]
	(0.034)	(0.021)	(0.029)	
English Z-score	0.063	0.045	0.174	[0.01, 0.20]
	(0.053)	(0.026)	(0.024)	
Math Z-score	0.042	-0.034	0.099	[-0.07, 0.13]
	(0.062)	(0.029)	(0.023)	

Notes: This table reports the upper and lower bounds of the lee bounding exercise for all intermediary outcomes and test scores. Standard errors associated with the bounds are included in parenthesis. 90% confidence intervals are reported in column 4.

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*

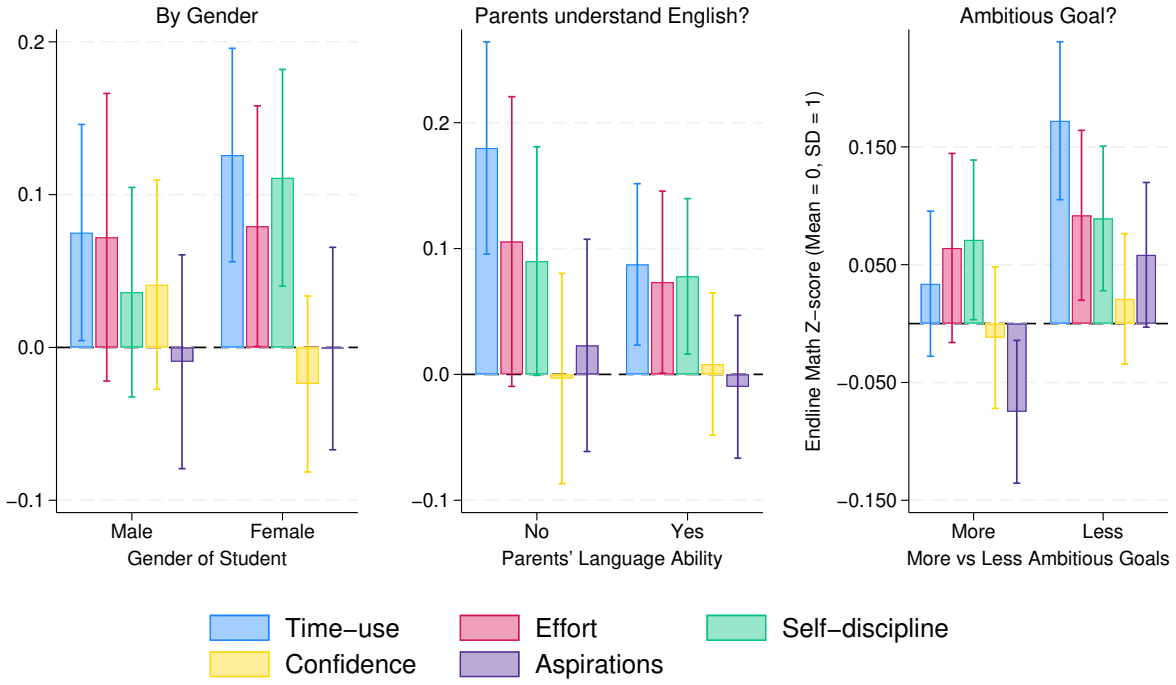
Figure A.1: Script for "Goal Setting" Schools - Part 2 of 2

	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.

Figure A.3: Heterogeneity in the Impact on Intermediary Outcomes: By Gender, Parents' English language ability and by the ambitious nature of the set goal



Notes: This figure shows the impact of goal-setting on intermediary outcomes (time-use, effort, self-discipline, confidence, and aspirations) split by gender of the student, parents' english language ability and the extent to which the set goals are ambitious. Baseline values of all dependent variables are included as controls. Standard errors are clustered at the level of school. Bars around the coefficient represent a 90% confidence interval.