Annenberg Institute
BROWN UNIVERSITY

EdWorkingPaper No. 25-1344

Learning to Work Towards Goals: A Sequential Evaluation of the Effect of Goal-Setting Course on Academic and Soft Skills

Anaya Dam
University of Oxford

Guthrie Gray-Lobe
University of Chicago

Michael Kremer
University of Chicago

Joost de Laat
Wageningen University

Karlijn Morsink
Utrecht University

This study sequentially evaluates a soft-skills course implemented in Ugandan and Kenyan primary schools that replaced academic review time with lessons on goal-setting and related skills as students prepared for high-stakes primary school-leaving exams. An exploratory evaluation in Uganda provided evidence of positive impacts on girls' test scores. A confirmatory evaluation in Kenya found that the course led to improvements in self-reported soft skills, especially among girls, although no gains in test scores. The study illustrates the utility of sequential evaluation, with exploratory analysis to identify promising hypotheses, followed by out-of-sample testing, as a tool to uncover heterogeneous effects.

VERSION: November 2025

Suggested citation: Dam, Anaya, Guthrie Gray-Lobe, Michael Kremer, Joost de Laat, and Karlijn Morsink. (2025). Learning to Work Towards Goals: A Sequential Evaluation of the Effect of Goal-Setting Course on Academic and Soft Skills. (EdWorkingPaper: 25-1344). Retrieved from Annenberg Institute at Brown University: https://doi.org/10.26300/51p5-g795

Learning to Work Towards Goals: A Sequential Evaluation of the Effect of a Goal-Setting Course on Academic and Soft Skills*

Anaya Dam[†] Guthrie Gray-Lobe[‡] Michael Kremer[§]
Joost de Laat[¶] Karlijn Morsink [‖]

This version: November 19, 2025

This study sequentially evaluates a soft-skills course implemented in Ugandan and Kenyan primary schools that replaced academic review time with lessons on goal-setting and related skills as students prepared for high-stakes primary school-leaving exams. An exploratory evaluation in Uganda provided evidence of positive impacts on girls' test scores. A confirmatory evaluation in Kenya found that the course led to improvements in self-reported soft skills, especially among girls, although no gains in test scores. The study illustrates the utility of sequential evaluation, with exploratory analysis to identify promising hypotheses, followed by out-of-sample testing, as a tool to uncover heterogeneous effects.

*We are grateful to Patricio Dalton, Giuseppe Sorrenti, Damon Clark, and Pablo Peña for comments. We thank audiences at the CSAE Conference 2024 and Advances in Field Experiments Conference 2024. The evaluation received support from the J-PAL Post-Primary Education Initiative, the International Growth Centre (IGC), and the Development Innovation Lab. This study reports results from an experiment designed and conducted in schools operated by NewGlobe. We thank Shannon May, Sean Geraghty, Tim Sullivan, and Clotilde de Maricourt for collaboration and valuable feedback. We thank Cynthia Ji for invaluable research assistance. IRB Approval was received from the Social and Behavioral Sciences (SBS) IRB of the University of Chicago (No IRB20-2181). The evaluation was registered at the American Economic Association's registry for randomized controlled trials (AEARCTR - 0010091). The registration includes a Preliminary Analysis and Plan For Subsequent Analysis (PAPSA) document developed while partially blinded to the data.

[†]University of Oxford, email: anaya.dam@bsg.ox.ac.uk
[‡]University of Chicago, email: graylobe@uchicago.edu
[§]University of Chicago, email: kremermr@uchicago.edu
[¶]Wageningen University, email: joost.delaat@wur.nl
[‖]Utrecht University, Wageningen University & Cornell University, email: k.morsink@uu.nl

1 Introduction

The ability to self-regulate learning may enable students to persist, adapt, and ultimately excel in school (e.g. Pintrich, 2000; Zimmerman, 2002). Self-regulated learning involves the capacity to define attainable goals, create plans to pursue them, and then monitor and adjust those plans as circumstances change (Winne and Hadwin, 1998; Zimmerman, 2002). Small-scale studies in psychology find that even brief interventions – such as teaching students how to set goals and providing a structured space to do so – can yield positive, and in some cases, large improvements in test performance (e.g. Chase et al., 2013; Morisano et al., 2010; van Lent and Souverijn, 2020). These studies suggest that the act of formulating a well-conceived goal and plan can serve as an internal commitment device, sharpening incentives for students to exert costly effort in the present to advance academically in the future. However, the effectiveness of goal setting as a self-regulatory tool depends on a broader set of skills and habits, from planning and persistence to reflection and adjustment. Treating goal-setting not as a one-off exercise, but as a course in itself – one that reinforces key lessons over time and helps students build enduring habits through repeated practice – may be important to help students excel academically through self-regulated effort.

We evaluate a large-scale and intensive goal-setting course implemented in primary schools in Uganda and Kenya. The course's objective was to prepare students for a critical juncture in their educational careers – national end-of-primary school exams, which determine secondary school transition. The goal-setting course replaced 25 sessions of revision class over one academic year (875 minutes of total lesson time). The course was implemented by teachers and mimicked the pedagogy of a curriculum for instruction on hard academic skills – with workbooks, teacher instruction, and self-guided practice time. In a given lesson, students were introduced to the goal-setting topic of that week, given time for independent goal setting, and encouraged to continue working on these goals outside of class.

We conduct a sequential evaluation where an initial exploratory study was carried out to

guide a second round, i.e., an out-of-sample evaluation. The exploratory study – conducted in Uganda – revealed some positive effects on national exam test scores. Students who received the goal-setting course scored 0.129 standard deviations higher on the Ugandan Primary School Leaving Exam (PLE), although the estimate was not statistically significant. Larger and nominally statistically significant impacts were observed for girls. These results motivated a confirmatory evaluation of the program in Kenya, where additional data on intermediate skills targeted by the intervention were collected. The second evaluation was guided by a pre-analysis plan that reflected the hypotheses that had received empirical support from the exploratory study.[1]

In Kenya, the goal-setting course led to improvements in self-reported soft skills related to goal-setting, with robust impacts among female students. While the intervention significantly impacted grit of all students, female students in particular show gains across a much broader range of soft-skill measures, including planning behaviors and patience. Despite finding positive effects on soft skills, the goal-setting course had no discernible significant effect on our pre-specified test score measures. The effect on the end-of-primary Kenya Certificate of Primary Education (KCPE) exam is 0.011 standard deviations, and the effect on the teacher-graded end-of-term assessment conducted shortly after the end of the program is -0.077 standard deviations. For neither outcome is there evidence of a larger effect for girls.

While the confirmatory evaluation in Kenya did not replicate the impacts on end-of-primary exams observed in Uganda, including for girls, the evidence generally suggests that spending school time teaching a goal-setting course does not negatively affect hard skills. In Kenya, we find no evidence of negative impacts on English, for which instruction was reduced to accommodate the goal-setting course. Ostensibly, the goal-setting course appears to increase soft skills without harming hard skills.

Furthermore, we note that the intervention did not affect students' expectations about

[1]The analysis plan outlined the primary objectives of the study as 1) assessing whether the course could improve academic outcomes, 2) whether the effects differ by gender, 3) the degree to which future expectations and soft skills mediate the effect, and 4) whether the effects depend on students' initial performance.

their performance or change their reported behaviors, such as study time.

Self-reported soft skills are strongly correlated with test scores, consistent with other related work (e.g Beattie et al., 2018; Duckworth et al., 2007; Heckman and Kautz, 2012; Kautz et al., 2014; Smithers et al., 2018). Motivated by this relationship, we use surrogate analysis to estimate the test score gains implied by the intervention's effects on soft skills (Athey et al., 2025; Prentice, 1989). Assuming that any impact on test scores operates through changes in soft skills, the implied aggregate effect on KCPE scores is 0.08 standard deviations, and 0.12 standard deviations for girls. These gains are modest—smaller than what this study, and most studies in the literature, would be powered to detect.

To contextualize the results, this study presents quantitative analyses of the literature on the effect of soft-skills interventions on academic test scores. We identified 33 experimental evaluations of soft-skills interventions, and meta-analyses. This analysis leads to two observations. Goal-setting interventions appear to be especially impactful, suggesting that spending time developing goal-setting skills could be a highly effective intervention. Second, small-scale studies of goal-setting find much larger effects while large-scale studies find small to no effects, potentially reflecting a higher likelihood of detecting false positives among small samples or signaling a loss of "voltage" (List, 2022).

This study makes several contributions. First, to our knowledge, this is the first experimental evaluation of a *course* in goal setting. It adds to the growing literature on the impact of setting goals on academic outcomes. The longest experimental evaluation of the effect of goal-setting is Morisano et al. (2010), which evaluates the impact of a 2.5 hour activity outside of the classroom. Dobronyi et al. (2019) and Oreopoulos and Petronijevic (2019) evaluate two hour activities. Evidence from these nudge experiments is mixed, with some studies showing very large impacts (e.g. Chase et al., 2013; Morisano et al., 2010), and others finding more modest impacts (e.g. Clark et al., 2020; Dobronyi et al., 2019; Islam et al., 2024). Still, the evidence base for goal-setting indicates potentially large impacts of teaching goal setting and related skills. We extend this literature by testing whether intensive training

to develop goal-setting *skills* yields more consistent gains, evaluating a course that teaches effective goal formation, strategies to overcome barriers, and repeated practice.

Second, our study contributes to our understanding of strategies for the scalability of soft-skills instruction in schools. The observed decline in efficacy of soft-skills interventions in education when studies are scaled could reflect span-of-control issues that arise when programs are implemented at larger scale, especially when expert research involvement is removed (List, 2022). The experimental results of this study represent the first experimental evaluation of a soft skills course deployed using structured pedagogy, a key tool for standardizing instructional quality and expanding organizational span of control. This tool may be especially important when bringing educational interventions to scale. A large literature demonstrates that structured pedagogy, which typically features teacher guides and complementary training and coaching to standardize how educational content is delivered, can be an effective tool for improving the quality of instruction in LMICs (Gray-Lobe et al., 2022; Piper et al., 2015, 2018a,b), and has been identified as a "Great Buy" in the most recent World Bank report on the Global Education Evidence Advisory Panel (GEEAP, 2023). The results of the study show that a goal-setting course deployed using structured pedagogy produced impacts on some self-reported soft skills, but did not produce gains in test scores.

This study illustrates the potential value of sequential exploratory and out-of-sample testing as an approach to testing hypotheses, adding to a growing literature on optimal evaluation (Azevedo et al., 2020; Banerjee et al., 2017) and identification of sub-group effects (Chernozhukov et al., 2020; Imai and Ratkovic, 2013; Wager and Athey, 2018). Researchers must allocate their time and effort to exploring hypotheses that are likely to yield results interesting to broad audiences. Research conventions, such as the norm of writing detailed pre-analysis plans, impose moderate fixed costs of time on researchers. Because exploratory research can be more affected by publication bias, strong PAP norms limit the scope for Type I error. However, when exploratory research is heavily discounted, useful empirical inquiries may not be pursued and entered into the literature. When researchers can sequence

exploratory and PAP-compliant research, the set of hypotheses that a researcher would pursue voluntarily can be expanded. Sequential evaluation can be especially valuable because researchers have private information about the quality of exploratory research. While not all research topics have the ability to conduct a low-cost exploratory study, the increasing involvement of researchers with organizations conducting A/B testing may mean that this approach could be broadly useful. As such, we view this study as closely related to other work on optimal evaluation strategies in A/B testing, such as Azevedo et al. (2020), which does not consider fixed costs of research and the possibility of sequential, out-of-sample testing.

The remainder of the paper is organized as follows: Section 2 provides context for the study, including meta-analytical results from a review of studies of the impact of soft-skills education on test scores, information on the school system in which the study took place, and the research design; Section 3 describes the experimental design; Section 4 describes the data used in the analysis of the experiments; Section 5 presents the results; and Section 6 discusses interpretation and concludes.

2 Background

2.1 Studies on the impact of soft-skills education on test scores

We reviewed 33 experimental evaluations of soft-skills interventions in education to contextualize the findings from Uganda and Kenya. Studies included programs aimed at using goal setting as well as those influencing mindset, working memory, social and emotional skills, and self-regulation. The studies are listed in Appendix Table A1.

Fifteen studies reported a single test score outcome. For studies with multiple outcomes, we applied the following decision rules to obtain a single point estimate: (i) if the intervention focused on a specific subject or exam, we used that assessment (1 study); (ii) where multiple outcomes qualified, we prioritized researcher-designed tests, followed by standardized exams

(e.g., national exam), then internal measures such as GPA or grades (3 studies); (iii) if assessments were administered at multiple points in time, we selected the measure closest to one year post-intervention (11 studies); and (iv) where multiple subject outcomes remained after the above criteria were applied, we constructed a precision-weighted average (8 studies). Authors were contacted to confirm estimates. Studies without responses are indicated in the table.

The impacts of soft-skills interventions on test scores tend to be substantial. Random effects meta-analysis indicates an effect of 0.13 standard deviations across all studies (Figure 1). Goal-setting interventions in the literature appear to be very light touch, 2.5 hours being the highest duration Morisano et al. (2010). Despite this, the average effect of the 10 goal-setting interventions on test scores is 0.15 standard deviations, compared to 0.11 standard deviations for other soft-skill interventions.

We see that the efficacy of soft-skills interventions tends to decline with the scale of the evaluation, as measured by the number of students in the treatment arm. Appendix Figure A1 shows a pattern of declining efficacy with scale that holds for the sub-sample of goal-setting studies and applies broadly to other soft skill interventions as well. In addition, all studies with fewer than 50 students have impacts on test scores above 0.2 standard deviations (Chase et al., 2013; Cook et al., 2014; Good et al., 2003). However, only two studies with more than 1,000 students have generated impacts close to 0.2 standard deviations (Alan et al., 2019; Porter et al., 2022).

The decline in efficacy of interventions with scale of the evaluation may reflect at least three factors. First, the downward slope may reflect a reduction in span-of-control with increasing scale (Lucas, 1978). Researchers have become increasingly concerned that small-scale, researcher-designed programs produce large positive effects that cannot be replicated at large scale, potentially because the span-of-control over the quality of implementation of the program (List, 2022). Second, these results may reflect the higher likelihood of overestimating standard errors in small samples (*M-error*), leading to an increased chance of detecting

false positives (Gelman and Carlin, 2014; Gelman et al., 2020). Third, these result may reflect publication bias if larger-scale evaluations are more likely to be published.[2] Concerns about publication bias have been raised in the literature on growth mindset interventions (Macnamara and Burgoyne, 2023). Appendix Figure A2 shows a funnel plot with the p-value from Egger's test of publication bias being 0.013 (Egger et al., 1997).

2.2 Intervention context and schools

The evaluations took place in NewGlobe's fee-charging schools in Kenya and Uganda, branded as Bridge Kenya and Bridge Uganda schools (henceforth: Bridge). NewGlobe is an international for-profit educational provider.[3] Bridge schools charge fees of around 100 USD per year – about 6-10 percent of GDP per capita in Kenya and Uganda. They are typically located near lower-income populations, such as in informal settlements or rural population centers.

Bridge schools use detailed tablet-based teacher guides with instruction that is standardized across schools. Guides are used for all academic subjects, and teachers receive specialized training on their use. Classroom-level lesson completion is monitored via the tablet, and supervisors conduct routine classroom observations. Gray-Lobe et al. (2022) find that this model is effective at teaching hard skills, showing that primary-school-aged Kenyan students in Bridge schools who get taught with the standardized teacher guides perform nearly a full grade level above those who attend other public and private schools in all academic subjects.

Bridge students tend to perform slightly above average on the KCPE. In 2022, the average

[2]The Cochrane Handbook notes that publication bias concerns may be greater for small studies (Higgins et al., 2022). There are several reasons small-scale studies may be more prone to publication bias. First, if small-scale studies lack the power to detect the impact of soft-skill intervention on test scores, then those small-scale studies that reject the null hypothesis will tend to have large, but biased effects (Gelman and Carlin, 2014). If the research community tends to prefer estimates that support a particular view, then small studies will tend to remain "in the file drawer" (Oyserman, 2023).

[3]At the time of the study, the firm operated private schools in Kenya, Uganda, Nigeria, and India, and provided services supporting public schools in Liberia, Nigeria, and Rwanda. The private schools in Kenya and Uganda have since been restructured into independent non-profits and continue to charge fees.

KCPE score for Bridge students was 263, compared to 250 nationally (Bridge International Academies Kenya, 2022). Among students for whom KCPE data were shared as part of this study, we see that they are less likely to appear in the bottom of the distribution, with 19 percent of students scoring below 200, compared to 24 percent nationally (Appendix Table A2). Bridge students are slightly less likely to be in the top of the distribution, with only 0.13 percent scoring above 400, compared to 0.77 percent nationally.

Bridge students spend more time at school and are more likely to report their teachers are teaching when they are in class. Gray-Lobe et al. (2022) find that upper primary Bridge students report spending about 40 minutes more in school on weekdays than their counterparts in public schools. Three-quarters of upper primary students in Bridge schools report attending school on Saturdays, compared to only one-quarter in public schools. Students and parents also report getting more individualized attention from teachers than those in public schools. Class sizes are smaller, with public schools having around 34 students per teacher and Bridge schools having 20. While the level of traditional educational inputs is generally higher in Bridge schools than in other public schools, Bridge schools' practices are similar to those of schools in wealthier countries, as well as those of many private schools in Kenya.

2.3 Intervention content

2.3.1 The goal-setting course

The goal-setting course materials were designed by the education innovation team at NewGlobe, consisting of pedagogists, teachers, and psychologists. While teachers received soft-copy versions of the structured teacher guides that directed the lesson for each session on tablets, students received the curriculum in hard-copy paper workbooks for them to fill out.[4] NewGlobe designed this curriculum to dedicate time to an in-school goal-setting subject aimed at improving students' performance in the end-of-primary exam – a critical juncture

[4]The workbook used in the Kenya study can be viewed here.

in the students' education trajectory. The goal-setting course is one of many strategies in Bridge schools to help students improve performance on high-stakes national end-of-primary exams (Gray-Lobe et al., 2025). The content was piloted and contextualized to be engaging for the age group.

In Kenya, the weekly goal-setting course replaced an English review session. Bridge Kenya's English instruction includes a daily English revision class, so the goal-setting intervention reduced time in English revision sessions by around 20 percent. In Uganda, the goal-setting course displaced review, but in that case, the subject was rotated weekly.

The course covered 25 lessons (See Appendix A.1 for the goal-setting course syllabus). The course took 875 minutes, longer than all except four of the soft-skills interventions identified in the review above. The lesson plan focused on helping students break down their envisioned future outcomes – such as becoming a doctor, teacher, pilot, or entrepreneur – into actionable short-term study and learning objectives. Students were taught how to identify their strengths and weaknesses in academic subjects and topics, and then chose specific tasks to complete in the present, such as revising content on that subject or topic, doing homework, and practicing mock exams. Students also learned to choose these objectives with the upcoming end-of-primary exam in mind, guiding them towards proximate milestones in achieving their longer-term future goals. The lesson was structured in two distinct parts: an instructional component led by the teacher, where the content was delivered and clarified, followed by an independent practice component where students worked on their self-set goals. At the end of each class, teachers checked in with the students and asked a selection of them to share their progress. Students were further encouraged to continue working on their goals and plans, for example during other classes, or while doing their homework. The key features of the content can be grouped into:

1. **Hierarchical goal structuring**: Lessons begin with defining goals and breaking down initial goals in the SMART framework (Lesson 1, 3 & 24). Students learn to set short-, medium- and long-term goals (Lesson 2, 7 & 9). They are asked to visualize the future

outcomes by reflecting on why achieving them matters for themselves (Lesson 8), and at regular intervals they rank and (re-)prioritize their short-term objectives to align with the future outcomes and progress (Lessons 10, 13, 14, 15).

2. **Ongoing evaluation and adjustment**: At regular intervals, assessments of progress towards the short-term objectives (Lessons 5, 11, 17, 22) and adjustment of goals (Lessons 6, 12, 18, 23) help students refine their objectives, as needed.

3. **Addressing challenges**: Some lessons address topics of procrastination (Lesson 4) and how to overcome obstacles and persevere despite setbacks (Lesson 25).

4. **Practical strategies for goal achievement**: Students are taught tools to manage time effectively and track progress, such as through providing a timetable and to-do checklist templates (Lesson 16, 19 & 21), as well as interactive goal ladders (Lesson 24). Students are also taught study tips (Lesson 20 & 25).

2.3.2 The revision class

In the comparison schools, students followed the status quo timetable. The goal-setting course displaced revision class for academic content. In Kenya, this time was specifically focused on English revision. In Uganda, this was revision of any academic subject.[5] The goal-setting intervention, thus, did not affect the introduction of new content in treatment schools. The revision class was highly structured – content and practice questions were provided to the teacher in the scripted teacher guide. Teachers led the class, provided students with practice questions, and students followed along in solving these with the teacher.

2.4 Lesson monitoring and intervention fidelity

NewGlobe's model of structured pedagogy – of which the structured teacher guides are a component – features frequent monitoring by supervisors and quality assurance staff (Gray-

[5]In Uganda, the national curriculum does not introduce new material in Grade 7.

Lobe et al., 2022), ensuring that teachers implement lessons as designed. School managers conduct regular check-ins to observe instruction and verify that teachers are following the guides. Lesson completion rates in the classrooms implementing the goal-setting course were similar to those doing the revision class, at about 80% on average (Appendix Figure A3).

3 Experimental design

3.1 Sample

The study takes place in Grade 7 classrooms. Grade 7 is the penultimate year of primary school in Kenya and the final year of primary school in Uganda. At the end of primary school, in both countries, students sit for the nationally administered and graded exam by the Ministry of Education: the Kenya Certificate of Primary Education and the Uganda Primary Leaving Exam. Both exams play a role in whether students can advance to secondary school and the types of secondary schools they can access.[6] Scores are also used in assigning public and private merit-based scholarships to pay for secondary school.

The study sample consisted of all active students in Term 1 who were enrolled in Grade 7 in Bridge schools.[7] Students who joined Bridge after the intervention started are excluded from the analysis to ensure that post-assignment changes in composition do not confound estimates. In Kenya, 1,963 students across 110 schools meet these criteria, while in Uganda, 327 students from 39 schools do.

[6] In Uganda, students falling below a cut-off are not eligible for Senior One. In Kenya, at the time of the study, there was no explicit cut-off; however, the public secondary school system is oversubscribed, so some students may score low enough to be ineligible for secondary school. A student who is not admitted to any public secondary school in Kenya can access private secondary school options, but these tend to be substantially more expensive than the public sector.

[7] The student roster contains many inactive students who register before the start of the academic year, but never end up showing up to school. We address this by restricting our sample to students who are 'active' in Term 1, i.e., close to the start of the intervention. In Kenya, we make this restriction by focusing on students who have non-zero attendance and have attended at least one exam in this term. In Uganda, where the end-of-term 1 'roster' of students already accounts for this, we use the students contained in this roster as the starting sample.

3.2 Randomization

Before the start of the academic year, schools were randomly assigned to one of two experimental arms. In Uganda, 18 out of 39 schools, and in Kenya, 55 out of 110 schools were randomly assigned to the goal-setting curriculum. In Uganda, randomization was done by blocking on four regional strata. To increase power, in Kenya, pairwise randomization was conducted within province by matching on the school-level average national exam score in the previous year.[8] The tablet-based lesson guides allow us to establish that all schools complied with their treatment assignment.

Timelines of the Uganda and Kenya studies are presented in Figures 2 and 3.

4 Data and descriptive statistics

4.1 Data

The study uses **anonymized administrative data** collected in Bridge schools as part of their business operations. Demographic data (e.g., gender, age, and years of enrollment in its schools), results from teacher-graded assessments, national end-of-primary exams, and data tracking graduates' post-primary outcomes were shared.

Teacher-graded assessments reflect performance on centrally-designed (at the country level) assessments. Assessments are conducted in all academic subjects each term. The Term 1 exams are conducted at baseline, Term 2 exams were conducted during the intervention. Term 3 exams were conducted just after the intervention ended. At the time of the study,

[8]Firstly, within a province, schools were rank-ordered based on the average total national exam test score for all students in the school in the previous academic year. There are three provinces – Coastal, Central, and Western. Secondly, within each province, based on the rank-ordered list, strata were created by taking two consecutive schools at a time. Two provinces had an odd number of schools, therefore, these provinces had one stratum each with three (instead of two) schools. Schools in these strata were selected from the center of the rank-order listing (i.e., schools at position: (N+1)/2 -1, (N+1)/2, (N+1)/2 +1). Finally, within each stratum, one school was randomized to treatment and one to the control group using Stata-assigned random numbers. Within the two strata with three schools, two out of the three schools were chosen, and one was randomized to treatment and one to control. The third school was re-randomized with a 50-50 unconditional probability to be in treatment or control, using Stata-assigned random numbers.

schools transmitted counts of correct responses for each student on each subject assessment.[9]

NewGlobe maintains records of primary school graduates' national end-of-primary exam. These are externally administered and graded exams designed and implemented by the respective Ministries of Education. In Uganda, this is the Primary Leaving Exam (PLE), and in Kenya, the Kenya Certificate of Primary Education (KCPE). In Uganda, students take the national exam just after the intervention, in Grade 7, and in Kenya, students take this exam one year after the intervention, in Grade 8. We focus on test scores on the following set of subjects that are tested in both the internal and external exams – Social Science, Science, Math, Kiswahili (Kenya only), and English. To make test scores comparable across studies and time periods, we standardize test scores using the respective subject's control group mean for each experiment.

For the Kenya sample, there are two additional sources of survey data. First, NewGlobe **surveyed teachers and parents on secondary school transition** of graduates.[10] The data contains information on whether the student enrolled in secondary school as well as the tier of school. This data was verified with official communication the school had received from the secondary schools.[11]

A second source of survey data in Kenya consists of a **student-administered multiple-choice survey** to gather measures of the students' soft skills, motivation, well-being, and expectations. Students participated in this survey just before they took their final Term 3

[9]Item-level data were not retained at the time of the study.

[10]The transition between schooling levels, such as primary to secondary school, is a time period where drop-outs are typically observed. Reasons for this could include, geographical constraints, e.g., the highest levels of secondary school are typically clustered closer to urban settlements or big cities necessitating longer travel times, financial constraints, e.g., schools may still charge ancillary fees such as for uniforms, books or boarding (if the student's home is far) or as students grow older they may face higher opportunity costs in the labor or marriage market.

[11]Two rounds of this survey were conducted. In round 1, for those students in the study sample who were registered for the KCPE at NewGlobe, data was collected using a survey filled out by the academy manager. For 384 students, the academy manager did not have this information. The data for these students was gathered from parents or primary caregivers using a phone survey. In round 2, for the students in the study sample who were not registered to take the KCPE at NewGlobe, the corresponding data were collected from parents or primary caregivers using a phone survey. It also contained additional questions on KCPE test scores of these students.

exam, and thus before participating in the high-stakes national exam.[12] The survey contains 40 multiple-choice questions, and students were given 45 minutes of in-class time to complete it. It contains questions on expectations for test scores, as well as soft skills, such as patience, grit, planning and mental experiencing. We construct standardized indices for these outcomes as indicated in our pre-analysis plan (See Appendix A.2).

4.2 Balance

In both experiments, students receiving the goal-setting course are broadly similar to those in the comparison group across a range of observable characteristics. Table 1 shows descriptive statistics and balance for student characteristics collected at baseline. We report the mean and standard deviation in the control group, the standard error for the test that the difference in means between treatment and control is zero, and the normalized differences. Joint tests of differences on observable characteristics in Table 1 yield p-values of 0.728 and 0.202 for Uganda and Kenya, respectively. Half of the students are female. On average, students are 11 years old and have been enrolled in Bridge schools for just under 2 years. An implication of these results is that many students attended another primary school before they arrived at Bridge.

4.3 Attrition

Table 2 shows follow-up rates for exams in Kenya and Uganda separately, and for secondary school transition in Kenya. The follow-up rate for the national exam test scores is 91% in Uganda and 82% in Kenya. Attrition on the national exam outcomes is primarily due to test-taking status and exits from Bridge because the data is only available for those who complete primary school at Bridge. For other tests, follow-up rates vary: in Kenya, 81% for the Term 2 exam; 67% for the Term 3 exam. In Uganda, follow-up rates are 83% for

[12] In 14 schools, due to issues with survey delivery, the survey needed to be administered at the start of the subsequent academic year.

the Term 2 exam and 87% for the Term 3 exam.[13] Attrition on these outcomes is driven primarily by absence on the day of the test and students exiting Bridge.[14] The follow-up rate is 77% for the secondary school outcomes. Attrition on the secondary school survey may be due to exits from Bridge and survey non-response.

Data from the self-reported soft-skills survey was only received for 42% of the study sample. The higher rate of attrition is likely due to the unfamiliarity of teachers and students with the process used to administer the survey. Teacher data entry error may have played a significant role, as the technology used to transmit item-level data had only recently been adopted.

While follow-up raises some risk that estimates are confounded by selection among the sample of students for whom data are observed, there is no positive evidence to support this concern. The follow-up rate does not differ significantly between the treatment and control groups in either Uganda or Kenya. Conditional on observing outcome data, students in both experimental conditions appear to be broadly similar. Appendix Table A3 reports balance conditional on students having data on national exam scores. Appendix Table A4 shows that even for the soft skills survey, for which attrition is high, students who received the goal-setting course were broadly similar to those receiving the comparison group.

5 Results

In this section, we first discuss the extent to which expectations, soft skills and motivation are correlated with test score performance, before presenting the effects of the goal-setting course on these outcomes. Next we present the effects of the curriculum on test scores and secondary school enrollment.

[13]The Term 3 exam in Uganda is the midterm exam. The final exam in the year is the national exam, and not the term 3 endterm exam.

[14]Absence on the day of the test can be due in part to students being excluded from class due to missed fee payments.

5.1 Do expectations, soft skills, and motivation predict academic success?

In this section, we first examine the correlation between expectations, soft skills, and motivation with academic outcomes. First, we study the expectations of students regarding their performance on the upcoming national KCPE exam and the teacher-graded end-of-year exam. We study **expectations** because expectation formation and goal setting are connected. Theories of goal setting in economics take the perspective that goals can function as reference points, against which decisions to expend effort are evaluated, keeping the psychological cost of not meeting the goal in mind. Therefore, as long as the cost of effort is less than the psychological cost of not meeting the goal, an individual should exert the required effort to achieve it (Cettolin et al., 2024; Clark et al., 2020; Hsiaw, 2013; Koch and Nafziger, 2011; McKenzie et al., 2021). While expectations may not always coincide with goals, since they can reflect beliefs about factors beyond the student's control, they offer valuable context for comparing how students perform relative to their own expectations.

Next, we study the following **soft skills**: 1) patience, 2) planning or scheduling to make (efficient) use of (scarce) time for educational activities, 3) persevering towards one's goals despite setbacks or "grit," and 4) clarity in imagining one's future or "mental experiencing" (See Appendix A.2 for outcome construction). We then explore whether differences in how students are **motivated** predict academic outcomes. Using items from Ryan and Connell (1989), we measure students' intrinsic and extrinsic motivation.

Students in the Kenya study reported high expectations. Figure 4 graphs expectations against actual performance on the KCPE among students in the comparison and goal-setting course groups separately. While there is a strong positive relationship between expectations and performance, students are far more likely to overestimate their future performance. While 26.32 percent report that they expect to score between 400 and 500, among those for whom expectations data is available, only three (0.36 percent) scored above 400.

Students hold similarly high expectations for performance on the teacher-graded end-of-year exam (Appendix Figure A4) and the likely secondary school tier that students expect to enroll in (Appendix Table A5). While 78.2 percent of students report expecting to enroll in the most selective national schools, only 4.6 percent actually do so.

In terms of **soft skills**, measures of academic success, such as test scores and transitioning to secondary school, are strongly predicted by student-reported soft skills, albeit less so than baseline hard skills. Table 3 reports pair-wise correlations between several of the key variables in the comparison group. Grit, for which we find the most robust correlations with academic achievement, has a correlation of 0.33 with the Term 3 score and 0.41 for the national exam (KCPE). The measure of students' use of effective planning strategies has a correlation of 0.24 and 0.28 with the Term 3 and national exam (KCPE), respectively. Patience and mental experiencing are also strongly correlated with test score outcomes. The relationship between soft and hard skills and secondary school enrollment is less clear. The correlation between grit and secondary school enrollment is 0.11, and that for mental experiencing is 0.14. Planning is positively correlated with secondary school enrollment but the relationship is not statistically significant. The smaller correlations may not be surprising if one considers that a common reason for students' inability to enroll in secondary school is the inability to pay.

Measures of **motivation** also appear to predict test scores. Students reporting greater intrinsic motivation do not perform better on tests of hard skills. External motivation, i.e., motivation driven by gains or reward, or motivation driven by a desire to avoid punishment imposed by others, and introjected motivation, i.e., motivation coming from internal pressure like guilt avoidance or pride, are negatively correlated with test scores. While identified motivation, i.e., motivation coming from the appreciation of the importance of an outcome, is strongly positively correlated. Measures of motivation also predict secondary school enrollment. Perhaps surprisingly, more intrinsically motivated students are less likely to enroll in secondary school. One interpretation of this relationship could be that intrinsi-

cally motivated students are less inclined to do tasks for which they have a distaste to reach a desirable outcome in the future. Indeed, we find a positive relationship between identified motivation and extrinsic motivation with secondary school enrollment. Children who are more outcomes-focused are more likely to enroll in secondary school.

Table 3 reports the correlation between daily homework time and academic outcomes. Notably, we find at most a small positive, yet statistically insignificant, relationship between reported homework time and all academic achievement outcomes. Students who are higher-performing at baseline do not report spending more time on homework. There is also some evidence that higher levels of grit, planning, and identified motivation are associated with more accurate predictions of future test score performance (Appendix Table A6).

5.2 Malleability of soft skills and motivation

We estimate the following linear model:

$$S_{ias} = \alpha + \beta T_{as} + \Gamma X_{ias} + \tau_s + \epsilon_{ias} \tag{1}$$

where the dependent variable S_{ias} represents a measure of soft skills or other survey response student i, in school a and in strata s, and $T_{as} \in \{0, 1\}$ indicates whether the school was randomly assigned to the goal-setting intervention or the comparison condition. In estimating Equation 1, we also include randomization fixed effects and controls for a vector of pre-assignment characteristics denoted by X_{ias}. These controls are not necessary for identification, but are included for additional precision and were prespecified in the analysis plan. The vector X_{ias} includes separate controls for all subject-level baseline test scores (with scores imputed to zero when missing) and a corresponding vector of indicators representing whether imputation for missingness took place. Standard errors are clustered at the stratum-level, following De Chaisemartin and Ramirez-Cuellar (2024).

The goal-setting course increased measures of grit by 0.346 standard deviations in aggre-

gate (Panel A of Table 4). Point estimates for all other soft skills are positive, but statistically insignificant. The point estimate for patience is marginally statistically significant with a p-value of 0.14. A joint F-test that the aggregate effect of the goal-setting intervention on all measures of soft skills is marginally insignificant (p-value = 0.110). The impact of the goal-setting course on self-reported soft skills was largest for girls. Panel B reports estimates of a model that builds on Equation 1 by including an interaction with an indicator for whether the student was female. In Panel B, the main effect of the soft-skill intervention reflects the impact on boys, the interaction estimates the difference in the effect for girls and boys, and the total effect on girls – included at the bottom of the table – is the sum of the main and interaction effect. The main effect is statistically significant only for grit for boys. For grit, planning, and patience, the effects for girls are larger than for boys. The estimated effects for girls on grit, planning, and patience are 0.444 (p-value = 0.001), 0.226 (p-value = 0.070), and 0.301 (p-value = 0.011), respectively. The joint test for the interaction effect has a p-value of 0.029. The p-value for the effect on girls is 0.002. The joint test for the effect on boys cannot reject the null with a p-value of 0.620.

In Table 5, there appears to be a slight shift in students' **motivation** due to the intervention, with students reporting less introjected regulation (a tendency to rely on guilt or pride for motivation) by 0.185 standard deviations. This effect appears to be partly explained by an increase in identified regulation, a tendency to focus on outcomes (an increase of 0.123), although the estimated effect is not statistically distinguishable from zero. The estimated aggregate effect on subjective well-being is small and insignificant. A joint F-test that the aggregate effect on all measures of motivation and subjective well-being yields a p-value of 0.715. We also estimate a difference in the effect of the program for girls, although we cannot rule out the hypothesis that the effect for both boys and girls is zero. The p-value from a joint test for the effect on boys (p-value = 0.446) and girls (p-value = 0.789) cannot be rejected.

The goal-setting course does not appear to have an effect on students' expectations

about future test score performance. We test whether the intervention had any effect on the students' expected National Exam KCPE score and measures of the distance between their KCPE score and their expectations. Appendix Table A7 reports these results.

In Appendix Table A8, we show that the course does not appear to influence students' self-reported time spent on homework. On average, students in the comparison schools report spending about 63 minutes on homework on a usual day, and those in goal-setting classrooms report about 1 minute less, an effect that is statistically indistinguishable from zero.

5.3 Effect on test scores

For analysis of the impact on test scores, the primary specification is based on analysis at the student-subject level. Specifically, we estimate the following linear model of the effect of the goal-setting intervention on the performance of students on subject-level assessments:

$$Y_{iasv} = \alpha + \beta T_{as} + \Gamma X_{iasv} + \tau_s + \phi_v + \epsilon_{iasv} \ . \tag{2}$$

The dependent variable, Y_{iasv}, is the subject-level test score for student i, in school a, in strata s and in subject v. We include in X_{iasv} controls for the baseline performance for each subject in Y_{iasv} and fixed effects for strata (τ_s) and subject (ϕ_v). The advantage of this specification is that it allows for pre-existing within-student variation in performance across subjects to be absorbed by regression controls.[15]

In Uganda, estimates of Equation 2, which served as a pilot for the Kenya evaluation, the effect on the Term 3 exam is -0.072 standard deviations, and the effect on the PLE is 0.129 standard deviations. In Uganda, point estimates indicate that girls' test scores may have been improved by the goal-setting course. The estimated effect of the goal-setting course for boys is -0.093 standard deviations. The difference in effect between girls and boys is estimated to be 0.409 standard deviations. The total effect for girls is estimated to be 0.317

[15] We find results are very similar when allowing the stratum fixed effects to vary across subjects.

standard deviations, a statistically significant effect (according to nominal asymptotic tests) at less than the 10 percent level. This result, which motivated the replication in Kenya, more than eliminates the gender gap in PLE performance estimated in the comparison group of -0.228 standard deviations (the coefficient on the female indicator).

In Kenya, estimates of Equation 2 indicate that, in aggregate, test scores are unaffected by the goal-setting course. The estimated effect on the Term 3 assessment is -0.077 SD (Table 6). The upper bound of the 95 percent confidence interval is 0.091 standard deviations. For National Exam (KCPE) scores, the estimated effect is 0.011 standard deviations, and the upper bound of the 95 percent confidence interval is 0.15 standard deviations. In Kenya, there is no evidence of heterogeneous effects for girls, for whom impacts on soft skills were larger. For Term 3 scores, the interaction effect with the female indicator is 0.007 standard deviations, and the total estimated effect on girls is -0.075. For the National KCPE exam the interaction is 0.012 standard deviations and the total effect for girls is 0.018 standard deviations.

Pooling results from Uganda and Kenya yields similar results to those observed in Kenya given the substantially larger scale of that evaluation (see Appendix Table A9).

5.4 Effect on secondary school enrollment

The goal-setting course is associated with a 2.8 percentage point increase in secondary school enrollment (Table 6).[16] While the estimated effect is not statistically significant, it is important to note that the magnitude reflects a substantial reduction in the rate at which students fail to transition to secondary school. In the comparison group, 87.3 percent of students enrolled in some secondary school. A 2.8 percentage point increase in the likelihood that students enroll in secondary school represents a 22 percent reduction in the fraction of students who do not transition.

We examine and find no impact on other outcomes related to secondary school enroll-

[16] To estimate this effect, we use an analogous specification to Equation 1.

ment, including whether applicants received a scholarship to offset costs of secondary school enrollment, and the type of secondary school that students enrolled in (Appendix Table A10).

5.5 Heterogeneous effects by baseline test scores

The impact of the course may have varied depending on students' initial performance. Meta-analyses of mindset interventions have found more robust positive impacts on test scores for lower-performing populations (Burnette et al., 2023; Tipton et al., 2023). Furthermore, a central component of the course is an introspective process of first understanding how they currently perform and then developing plans to reach their goals accordingly.

To explore this question, we estimate models that interact the goal-setting course with baseline performance. We create two variables for the baseline test score. For the first variable, we recenter the baseline test score to have a value of zero at the 25th percentile so that the main effect reflects the impact of a lower-performing student. For the second variable we create an indicator for whether the student is above or below the average of the test score distribution. Table 7 reports the results.[17]

In Kenya and Uganda, there is little evidence that the effect of the course depended on students' baseline test score performance. Estimates of the interaction effect are not statistically significant, regardless of the interaction variable used.

We do find some evidence of heterogeneity in the effect on soft skills, with significantly larger effects on self-reported patience for students with a higher baseline test score performance as compared to students with a lower baseline test score performance (Appendix Table A12).

[17] Appendix Table A11 reports heterogeneous reports analogous results for girls only.

5.6 Effect on displaced subject test scores

In Kenya, we find no evidence of negative effects on English scores, the subject for which instruction time was reduced to accommodate the goal-setting course. Appendix Table A13 and A14 reports effects by subject in Kenya.[18] The estimated effect on the English KCPE score is 0.09 standard deviations. For the teacher-graded end-of-year assessment, the effect on English is -0.11. In neither case are the estimated effects statistically distinguishable from zero.

5.7 Effects on gendered beliefs about academic performance

The soft-skills survey also asked students about their beliefs about whether girls or boys tend to perform better in math and English, and whether it was more important for girls to attend university. These questions were motivated by the findings in Uganda that showed improvements in girls' test scores. The results from Uganda indicated that girls were improving most in math and science, fields where they tended to lag behind boys. In Bridge Kenya, the gap between girls and boys in math tends to be smaller. Girls underperform boys in math by 0.11 standard deviations on the teacher-graded assessments (p-value = 0.036) and by 0.06 standard deviations on the KCPE (p-value = 0.222). Meanwhile, on language subjects, girls tend to outperform boys (0.14 standard deviations on teacher-graded assessments and 0.19 on the KCPE).

Even though the intervention did not affect observed academic performance, it appears to have influenced how students perceive the relative performance of girls and boys on academic subjects. Nearly 18 percent of students in the comparison group indicated that they believe that boys are better than girls at math (Table 8). This fraction was reduced by over ten percentage points in the schools receiving the soft-skills intervention. The fraction of boys reporting who believe that boys are better at math fell by 14.1 percentage points (Column

[18] For completeness, analogous results for Uganda are reported.

1). The reduction in gendered beliefs about math performance is offset by a 13.7 percentage point increase in the fraction of girls who indicate that girls tend to be better at math (Column 2).

5.8 Surrogate Index

If the course's impacts on academic skills are mediated by the soft skills that it teaches, what impacts would be expected given the impacts on the mediators? We address this question by constructing a surrogate index for academic outcomes (test scores and secondary school enrollment) using soft skills. To construct the surrogate index, we obtain a prediction of Y_i depending on a vector of soft skills and motivation S_i and pre-intervention test scores X_i. Specifically, we estimate the following linear model in the comparison group:

$$Y_i = S_i'\Gamma + X_i'\Omega + \varepsilon_i \ . \tag{3}$$

The surrogate index is the prediction from the model, $\hat{Y}_i$, representing the expected academic outcome given intermediate skills and pre-intervention academic skills. We then estimate the impact of the goal-setting course on the surrogate index.

$$\hat{Y}_i = T_j\delta + X_i'\Psi + \eta_{ij} \ . \tag{4}$$

The estimate, $\hat{\delta}$, represents the change in the expected test score given intermediate skills and pre-intervention academic skills under the assumption that the impact of the course on test scores is negligible after accounting for its impact on characteristics in S_i. This assumption would be violated if there were a direct effect of the course on test scores. For example, if the goal-setting course affects test scores through the commitment effect of setting goals, then this approach could understate the impact on test scores. To account for sampling variation in the first step, standard errors are estimated using a two-stage bootstrap, resampling individuals within schools first, and then resampling academies within strata.

The results from the surrogate index analysis indicate that, under the surrogacy assumption that all improvements in test scores are mediated by observed intermediate surrogates, the impacts on test scores would be too small to detect, given the precision of this study. In aggregate, the goal-setting course's effect on soft skills would be associated with a 0.050 standard deviation increase in teacher-graded term 3 assessments, a 0.078 standard deviation increase in KCPE scores, and a 1.1 percentage point increase in the likelihood of enrolling in secondary school (Table 9).[19] These effects are contained within the 95 percent confidence intervals of the aggregate effects reported in Table 6. We also use the bootstrap procedure to conduct inference on the difference between the impact on the observed outcome and the corresponding surrogate effect. For none of the outcomes is it possible to reject the null hypothesis that the effects are equal. Even for girls, for whom we observed gains across a broader range of soft skills, the impact on the surrogate index indicates only a 0.082 standard deviation increase in the teacher-graded term 3 score, and a 0.116 standard deviation increase in KCPE scores. Again, all of these estimates are well within the 95% confidence interval obtained from the analogous specification using the academic outcomes directly, and the p-values from the tests that the effects on the surrogate are the same as those on the outcomes are above conventional levels of statistical significance.

6 Conclusion

This study experimentally evaluated an intensive goal-setting course in Uganda and Kenya designed to help students develop and follow through on academic goals. An exploratory evaluation in Uganda found modest, but statistically insignificant, positive effects on the national primary school leaving exam, with large marginally significant gains for girls. A confirmatory study in Kenya found that students who received the course reported higher soft-skills, especially grit. The impact on reported soft skills of girls was larger and more

[19]Note that the standard errors for the impacts on test scores are smaller than those observed using the actual test scores as outcomes. This improvement in precision comes from the surrogacy assumptions, under which the test score includes variation unrelated to the surrogates (Athey et al., 2025).

robust, including positive effects on grit, planning, and patience. Despite soft skills gains, we find no statistically significant effects on test scores in Kenya.

The evaluation suggests that impacts on academic skills from an intensive goal-setting course are likely modest in aggregate. Surrogate analysis suggests that the observed soft-skill improvements could translate into between 0.08 and 0.12 standard deviations for girls. Our study stands in contrast to a large literature on the impacts of low-intensity and small scale goal-setting interventions. We are able to rule out aggregate effects in the range of the random effects estimate of 0.17 for all experimentally evaluated goal-setting interventions.

The impact of spending time teaching goal-setting may be different in other academic environments. The course was an effort to help students develop skills needed for self-regulated study. However, the study environment is one that already affords students high levels of external support that may not be present in many educational settings – with longer days, greater test preparation and external support, similar to that observed in charter schools (Angrist et al., 2013; Fryer, 2014; Thernstrom and Thernstrom, 2003). If external support substitutes for students' self-regulated study, this may limit the scope for teaching goal-setting to improve test scores. In other educational environments with less external support, the trade-offs between soft-skill and hard-skill instruction may differ. More generally, this study illustrates both the promise and the limitations of structured pedagogy for delivering soft-skill curricula at scale.

Remarkably, time spent on goal-setting does not appear to harm academic skills. If hard-skill instruction were more effective at producing academic outcomes than soft-skill instruction, displacement should reduce performance. Notably, we find no evidence of negative impacts on English in Kenya, despite the program displacing a disproportionate share of review time. Overall, results suggest that, within Bridge schools, teaching soft skills is about as productive for academic outcomes as teaching hard skills, while also raising (self-reported) soft skills.

Improvements in soft skills may provide benefits to students independent of hard skills,

some of which might only be realized later in their lives (e.g. Booker et al., 2011; Borghans et al., 2008; Deming, 2009, 2011; Gray-Lobe et al., 2023; Heckman et al., 2013; Heckman and Rubinstein, 2001; Sorrenti et al., 2024). While the present study is unable to assess long-term impacts of the goal-setting course, in Kenya, students taught soft skills were 2.8 percentage points more likely to enroll in secondary school. While statistically insignificant at conventional levels (p-value 0.244), this effect would represent a 25 percent reduction in the share of students who do not transition. Hence, we caution against inferring from our study that the gains in soft-skills from the goal-setting course have no meaningful impacts on education, labor market or later-life outcomes, which we were not able to measure.

References

Alan, S., Boneva, T., and Ertac, S. (2019). Ever failed, try again, succeed better: Results from a randomized educational intervention on grit. *Quarterly Journal of Economics*, 134(3):1121–1162.

Angrist, J. D., Pathak, P. A., and Walters, C. R. (2013). Explaining charter school effectiveness. *American Economic Journal: Applied Economics*, 5(4):1–27.

Aronson, J., Fried, C. B., and Good, C. (2002). Reducing the effects of stereotype threat on african american college students by shaping theories of intelligence. *Journal of Experimental Social Psychology*, 38(2):113–125.

Ashraf, N., Bryan, G., Delfino, A., Holmes, E. A., Iacovone, L., Meyer, C. J., and Pople, A. (2024). Learning to see a world of opportunities. *Unpublished Manuscript*. Available at: https://navaashraf.com/wp-content/uploads/2024/09/learning_to_see_a_world_of_opportunities-1.pdf.

Athey, S. A., Chetty, R., Imbens, G. W., and Kang, H. (2025). The surrogate index: Combining short-term proxies to estimate long-term treatment effects more rapidly and precisely. *Review of Economic Studies*. Forthcoming.

Azevedo, E. M., Deng, A., Montiel Olea, J. L., Rao, J., and Weyl, E. G. (2020). A/B testing with fat tails. *Journal of Political Economy*, 128(12):4614–000.

Banerjee, A., Chassang, S., and Snowberg, E. (2017). Decision theoretic approaches to experiment design and external validity. In Blattman, C. and Niehaus, P., editors, *Handbook of Economic Field Experiments, Volume 1*, pages 73–132. North-Holland, Amsterdam.

Beattie, G., Laliberté, J.-W. P., and Oreopoulos, P. (2018). Thrivers and divers: Using non-academic measures to predict college success and failure. *Economics of Education Review*, 62:170–182.

Berger, E. M., Fehr, E., Hermes, H., Schunk, D., and Winkel, K. (2025). The impact of working-memory training on children's cognitive and noncognitive skills. *Journal of Political Economy*, 133(2):492–521. Publisher: The University of Chicago Press.

Bettinger, E., Ludvigsen, S., Rege, M., Solli, I. F., and Yeager, D. (2018). Increasing perseverance in math: Evidence from a field experiment in Norway. *Journal of Economic Behavior & Organization*, 146:1–15.

Blackwell, L. S., Trzesniewski, K. H., and Dweck, C. S. (2007). Implicit theories of intelligence predict achievement across an adolescent transition: A longitudinal study and an intervention. *Child Development*, 78(1):246–263.

Booker, K., Sass, T. R., Gill, B., and Zimmer, R. (2011). The effects of charter high schools on educational attainment. *Journal of Labor Economics*, 29(2):377–415.

Borghans, L., Ter Weel, B., and Weinberg, B. A. (2008). Interpersonal styles and labor market outcomes. *Journal of Human resources*, 43(4):815–858.

Bridge International Academies Kenya (2022). National exam results: Year 8 (2022). https://bridge.ac.ke/kcpe-success/kcpe-results/national-exam-results-year-8-2022/. Accessed: 2025-08-08.

Broda, M., Yun, J., Schneider, B., Yeager, D. S., Walton, G. M., and Diemer, M. (2018). Reducing inequality in academic success for incoming college students: A randomized trial of growth mindset and belonging interventions. *Journal of Research on Educational Effectiveness*, 11(3):317–338.

Burnette, J. L., Billingsley, J., Banks, G. C., Knouse, L. E., Hoyt, C. L., Pollack, J. M., and Simon, S. (2023). A systematic review and meta-analysis of growth mindset interventions: For whom, how, and why might such interventions work? *Psychological Bulletin*, 149(3–4):174–205.

Cettolin, E., Cole, K., and Dalton, P. S. (2024). Improving workers' performance in small firms: A randomized experiment on goal setting in Ghana. *Review of Economics and Statistics*, pages 1–45.

Chase, J. A., Houmanfar, R., Hayes, S. C., Ward, T. A., Vilardaga, J. P., and Follette, V. (2013). Values are not just goals: Online ACT-based values training adds to goal setting in improving undergraduate college student performance. *Journal of Contextual Behavioral Science*, 2(3-4):79–84.

Chernozhukov, V., Demirer, M., Duflo, E., and Fernández-Val, I. (2020). Generic machine learning inference on heterogeneous treatment effects in randomized experiments. *American Economic Association: Papers and Proceedings*, 110:43–48.

Clark, D., Gill, D., Prowse, V., and Rush, M. (2020). Using goals to motivate college students: Theory and evidence from field experiments. *Review of Economics and Statistics*, 102(4):648–663.

Cook, P. J., Dodge, K., Farkas, G., Roland G. Fryer, J., Guryan, J., Ludwig, J., Mayer, S., Pollack, H., and Steinberg, L. (2014). The surprising efficacy of academic and behavioral intervention with disadvantaged youth: Results from a randomized experiment in Chicago. *NBER Working Paper No. 19862.* National Bureau of Economic Research.

De Chaisemartin, C. and Ramirez-Cuellar, J. (2024). At what level should one cluster standard errors in paired and small-strata experiments? *American Economic Journal: Applied Economics*, 16(1):193–212.

Dee, T. S. (2015). Social identity and achievement gaps: Evidence from an affirmation intervention. *Journal of Research on Educational Effectiveness*, 8(2):149–168.

Deming, D. (2009). Early childhood intervention and life-cycle skill development: Evidence from Head Start. *American Economic Journal: Applied Economics*, 1(3):111–134.

Deming, D. J. (2011). Better schools, less crime? *Quarterly Journal of Economics*, 126(4):2063–2115.

Dobronyi, C. R., Oreopoulos, P., and 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.

Duckworth, A., Peterson, C., Matthews, M., and Kelly, D. (2007). Grit: Perseverance and passion for long-term goals. *Journal of Personality and Social Psychology*, 92:1087–101.

Egger, M., Smith, G. D., Schneider, M., and Minder, C. (1997). Bias in meta-analysis detected by a simple, graphical test. *BMJ*, 315(7109):629–634.

Falk, A., Becker, A., Dohmen, T., Huffman, D., and Sunde, U. (2023). The preference survey module: A validated instrument for measuring risk, time, and social preferences. *Management Science*, 69(4):1935–1950.

Fryer, Roland G., J. (2014). Injecting charter school best practices into traditional public schools: Evidence from field experiments. *Quarterly Journal of Economics*, 129(3):1355–1407.

Ganimian, A. J. (2020). Growth-mindset interventions at scale: Experimental evidence from Argentina. *Educational Evaluation and Policy Analysis*, 42(3):417–438.

GEEAP (2023). *Cost Effective Approaches to Improve Global Learning*. Global Education Evidence Advisory Panel.

Gelman, A. and Carlin, J. (2014). Beyond power calculations: Assessing type s (sign) and type m (magnitude) errors. *Perspectives on psychological science*, 9(6):641–651.

Gelman, A., Skardhamar, T., and Aaltonen, M. (2020). Type m error might explain weisburd's paradox. *Journal of Quantitative Criminology*, 36:295–304.

Good, C., Aronson, J., and Inzlicht, M. (2003). Improving adolescents' standardized test performance: An intervention to reduce the effects of stereotype threat. *Journal of Applied Developmental Psychology*, 24(6):645–662.

Gray-Lobe, G., Jain, R., Kremer, M., de Laat, J., Schöpfer, P., and Scurlock, B. (2025). Does extra exam practice raise test scores? Experimental evidence from Kenya and Nigeria. *Unpublished Manuscript*.

Gray-Lobe, G., Keats, A., Kremer, M., Mbiti, I., and Ozier, O. W. (2022). Can education be standardized? Evidence from Kenya. *BFI Working Paper No. 2022-68*. Becker Friedman Institute for Economics, University of Chicago.

Gray-Lobe, G., Pathak, P. A., and Walters, C. R. (2023). The long-term effects of universal preschool in Boston. *The Quarterly Journal of Economics*, 138(1):363–411.

Heckman, J., Pinto, R., and Savelyev, P. (2013). Understanding the mechanisms through which an influential early childhood program boosted adult outcomes. *American Economic Review*, 103(6):2052–2086.

Heckman, J. J. and Kautz, T. (2012). Hard evidence on soft skills. *Labour Economics*, 19(4):451–464.

Heckman, J. J. and Rubinstein, Y. (2001). The importance of noncognitive skills: Lessons from the GED testing program. *American Economic Review*, 91(2):145–149.

Higgins, J. P., Thomas, J., Chandler, J., Cumpston, M., Li, T., Page, M. J., and Welch, V. A. e. (2022). *Cochrane Handbook for Systematic Reviews of Interventions, 2nd Edition.*

Hsiaw, A. (2013). Goal-setting and self-control. *Journal of Economic Theory*, 148(2):601–626.

Huillery, E., Bouguen, A., Charpentier, A., Algan, Y., and Chevallier, C. (2025). The Role of Mindset in Education : A Large-Scale Field Experiment in Disadvantaged Schools*. *The Economic Journal*, page ueaf015.

Hulleman, C. S. and Harackiewicz, J. M. (2009). Promoting interest and performance in high school science classes. *Science*, 326(5958):1410–1412.

Humphrey, N., Barlow, A., Wigelsworth, M., Lendrum, A., Pert, K., Joyce, C., and Turner, A. (2015). Promoting alternative thinking strategies (PATHS): Evaluation report and executive summary. *Evaluation Report No. ED581278.* Education Endowment Foundation.

Imai, K. and Ratkovic, M. (2013). Estimating treatment effect heterogeneity in randomized program evaluation. *Annals of Applied Statistics*, 7(1):443–470.

Islam, A., Kwon, S., Masood, E., Prakash, N., Sabarwal, S., and Saraswat, D. (2024). All pain and no gain: When goal setting leads to more effort but no gains in test scores. *Economics of Education Review*, 103(C).

Kautz, T., Heckman, J. J., Diris, R., ter Weel, B., and Borghans, L. (2014). Fostering and measuring skills: Improving cognitive and non-cognitive skills to promote lifetime success. *NBER Working Paper No. 20749.* National Bureau of Economic Research.

Koch, A. K. and Nafziger, J. (2011). Self-regulation through goal setting. *Scandinavian Journal of Economics*, 113(1):212–227.

List, J. (2022). *The Voltage Effect.* Penguin Books Limited.

Lucas, R. E. (1978). On the size distribution of business firms. *The Bell Journal of Economics*, 9(2):508–523.

Macnamara, B. N. and Burgoyne, A. P. (2023). Do growth mindset interventions impact students' academic achievement? A systematic review and meta-analysis with recommendations for best practices. *Psychological Bulletin*, 149(3-4):133.

McKenzie, D., Mohpal, A., and Yang, D. (2021). Aspirations and financial decisions: Experimental evidence from the philippines. Working Paper 28607, National Bureau of Economic Research.

Morisano, D., Hirsh, J. B., Peterson, J. B., Pihl, R. O., and Shore, B. M. (2010). Setting, elaborating, and reflecting on personal goals improves academic performance. *Journal of Applied Psychology*, 95(2):255–264.

Oreopoulos, P. and Petronijevic, U. (2019). The remarkable unresponsiveness of college students to nudging and what we can learn from it. *NBER Working Paper No. 26059.* National Bureau of Economic Research.

Outes-Leon, I., Sánchez, A., and Vakis, R. (2020). The power of believing you can get smarter: The impact of a growth-mindset intervention on academic achievement in Peru. *World Bank Policy Research Working Paper No. 9141* World Bank.

Oyserman, D. (2023). Culturally fluent theories, metascience, and scientific progress: A case example: Commentary on Macnamara and Burgoyne (2023) and Burnette et al. (2023). *Psychological Bulletin*, 149(3–4):220–228. Commentary; published online Oct 2022.

Oyserman, D., Bybee, D., and Terry, K. (2006). Possible selves and academic outcomes: How and when possible selves impel action. *Journal of Personality and Social Psychology*, 91(1):188.

Pintrich, P. R. (2000). Chapter 14 - the role of goal orientation in self-regulated learning. In Boekaerts, M., Pintrich, P. R., and Zeidner, M., editors, *Handbook of Self-Regulation*, pages 451–502. Academic Press, San Diego.

Piper, B., Jepkemei, E., Kwayumba, D., and Kibukho, K. (2015). Kenya's ICT policy in practice: The effectiveness of tablets and e-readers in improving student outcomes. *Forum for International Research in Education*, 2(1):3–18.

Piper, B., Sitabkhan, Y., Mejia, J., and Betts, K. (2018a). Effectiveness of teachers' guides in the Global South: Scripting, learning outcomes, and classroom utilization. *RTI Press - Occasional Paper.*

Piper, B., Zuilkowski, S. S., Dubeck, M., Jepkemei, E., and King, S. J. (2018b). Identifying the essential ingredients to literacy and numeracy improvement: Teacher professional development and coaching, student textbooks, and structured teachers' guides. *World Development*, 106:324–336.

Porter, T., Catalán Molina, D., Cimpian, A., Roberts, S., Fredericks, A., Blackwell, L. S., and Trzesniewski, K. (2022). Growth-mindset intervention delivered by teachers boosts achievement in early adolescence. *Psychological Science*, 33(7):1086–1096.

Prentice, R. L. (1989). Surrogate endpoints in clinical trials: Definition and operational criteria. *Statistics in Medicine*, 8(4):431–440.

Ryan, R. M. and Connell, J. P. (1989). Perceived locus of causality and internalization: Examining reasons for acting in two domains. *Journal of Personality and Social Psychology*, 57(5):749–761.

Salmela-Aro, K., Kiuru, N., Leskinen, E., and Nurmi, J.-E. (2009). School burnout inventory (SBI) reliability and validity. *European Journal of Psychological Assessment*, 25(1):48–57.

Samek, A., Gray, A., Datar, A., and Nicosia, N. (2021). Adolescent time and risk preferences: Measurement, determinants and field consequences. *Journal of Economic Behavior & Organization*, 184:460–488.

Schunk, D., Berger, E. M., Hermes, H., Winkel, K., and Fehr, E. (2022). Teaching self-regulation. *Nature Human Behaviour*, 6(12):1680–1690.

Smithers, L. G., Sawyer, A. C., Chittleborough, C. R., Davies, N. M., Smith, G. D., and Lynch, J. W. (2018). A systematic review and meta-analysis of effects of early life non-cognitive skills on academic, psychosocial, cognitive and health outcomes. *Nature Human Behaviour*, 2(11):867–880.

Sorrenti, G., Zölitz, U., Ribeaud, D., and Eisner, M. (2024). The causal impact of socio-emotional skills training on educational success. *Review of Economic Studies*, page rdae018.

Strathman, A., Gleicher, F., Boninger, D. S., and Edwards, C. S. (1994). The consideration of future consequences: Weighing immediate and distant outcomes of behavior. *Journal of Personality and Social Psychology*, 66(4):742.

Thernstrom, A. and Thernstrom, S. (2003). *No Excuses: Closing the Racial Gap in Learning*. Simon & Schuster, New York.

Tipton, E., Bryan, C., Murray, J., McDaniel, M. A., Schneider, B., and Yeager, D. S. (2023). Why meta-analyses of growth mindset and other interventions should follow best practices for examining heterogeneity: Commentary on Macnamara and Burgoyne (2023) and Burnette et al.(2023). *Psychological Bulletin*, 149(3–4):229–241.

van Lent, M. (2019). Goal setting, information, and goal revision: A field experiment. *German Economic Review*, 20(4):949–972.

van Lent, M. and Souverijn, M. (2020). Goal setting and raising the bar: A field experiment. *Journal of Behavioral and Experimental Economics*, 87:101570.

Wager, S. and Athey, S. (2018). Estimation and inference of heterogeneous treatment effects using random forests. *Journal of the American Statistical Association*, 113(523):1228–1242.

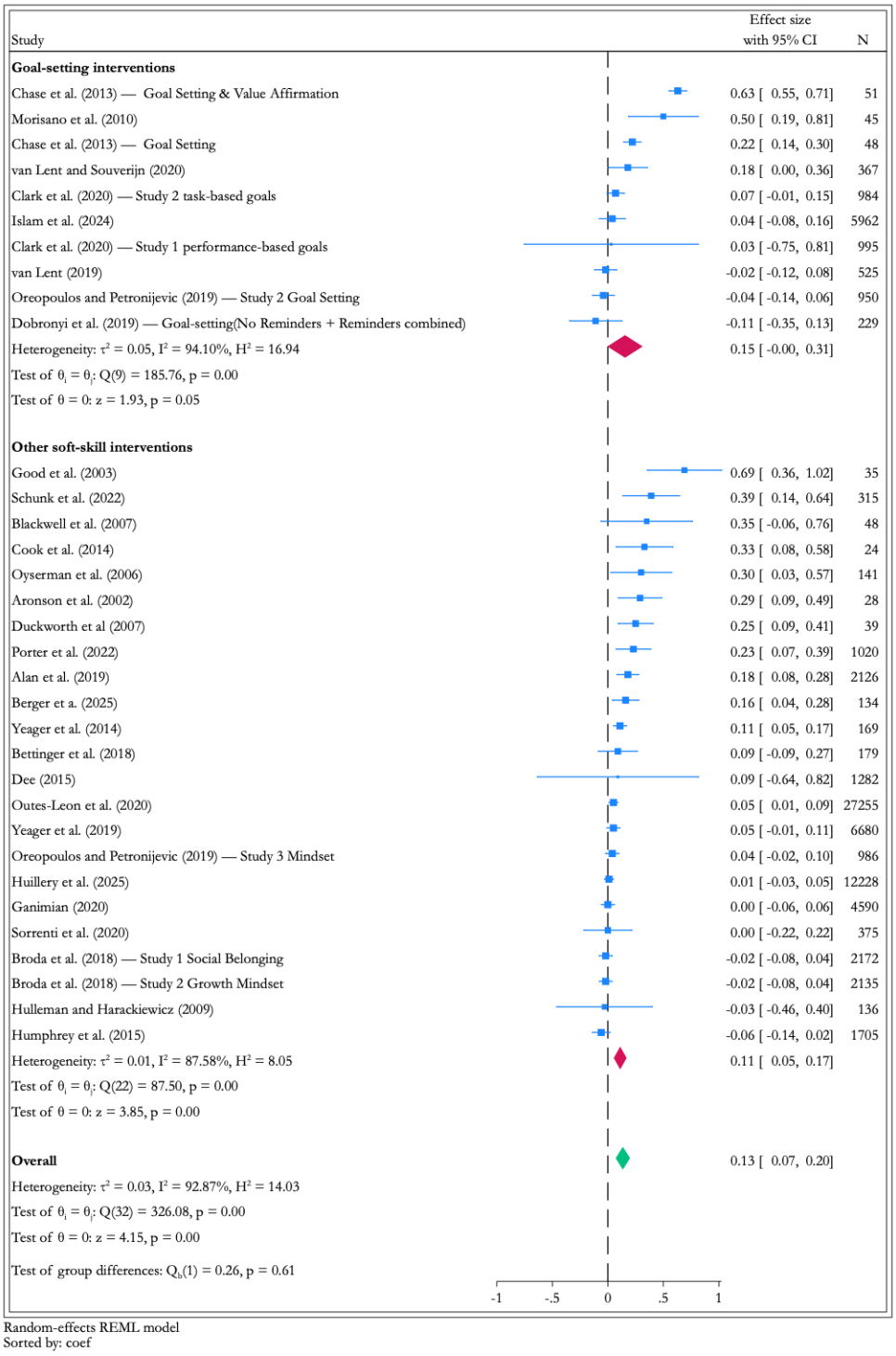
Winne, P. H. and Hadwin, A. F. (1998). Studying as self-regulated learning. In Hacker, D. J., Dunlosky, J., and Graesser, A. C., editors, *Metacognition in Educational Theory and Practice*, pages 277–304. Lawrence Erlbaum Associates, Mahwah, NJ.

Yeager, D. S., Hanselman, P., Walton, G. M., Murray, J. S., Crosnoe, R., Muller, C., Tipton, E., Schneider, B., Hulleman, C. S., Hinojosa, C. P., et al. (2019). A national experiment reveals where a growth mindset improves achievement. *Nature*, 573(7774):364–369.

Yeager, D. S., Henderson, M. D., Paunesku, D., Walton, G. M., D'Mello, S., Spitzer, B. J., and Duckworth, A. L. (2014). Boring but important: A self-transcendent purpose for learning fosters academic self-regulation. *Journal of Personality and Social Psychology*, 107(4):559.

Zimmerman, B. J. (2002). Becoming a self-regulated learner: An overview. *Theory Into Practice*, 41(2):64–70.

Figure 1: Forest plot of randomized evaluations of soft-skill interventions in education



Notes: The figure presents a forest plot of the 33 randomized evaluations of soft skills interventions in education, of which 10 evaluate a goal-setting intervention. Impacts on test scores in standard deviations and corresponding 95% confidence interval is presented. Studies are sorted by impact sizes. Additional details are in Table A1.

Figure 2: Timeline: Uganda

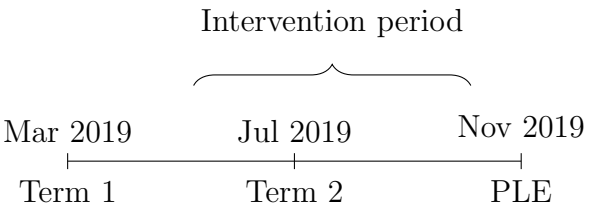


Figure 3: Timeline: Kenya

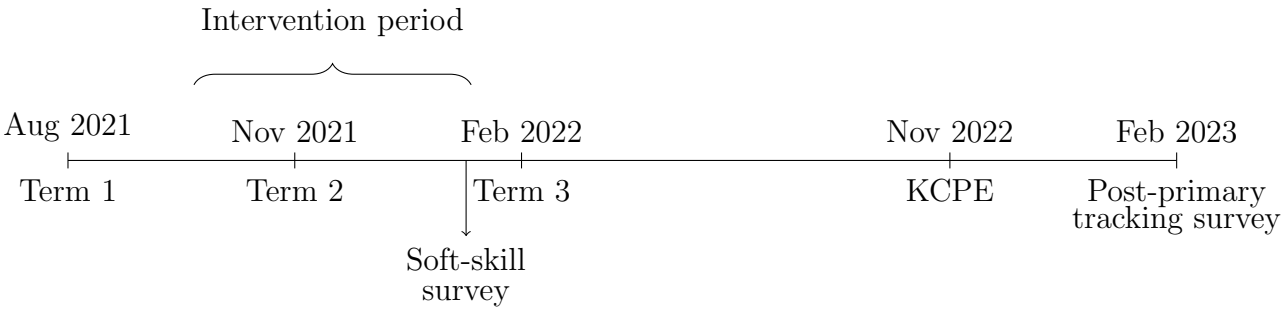
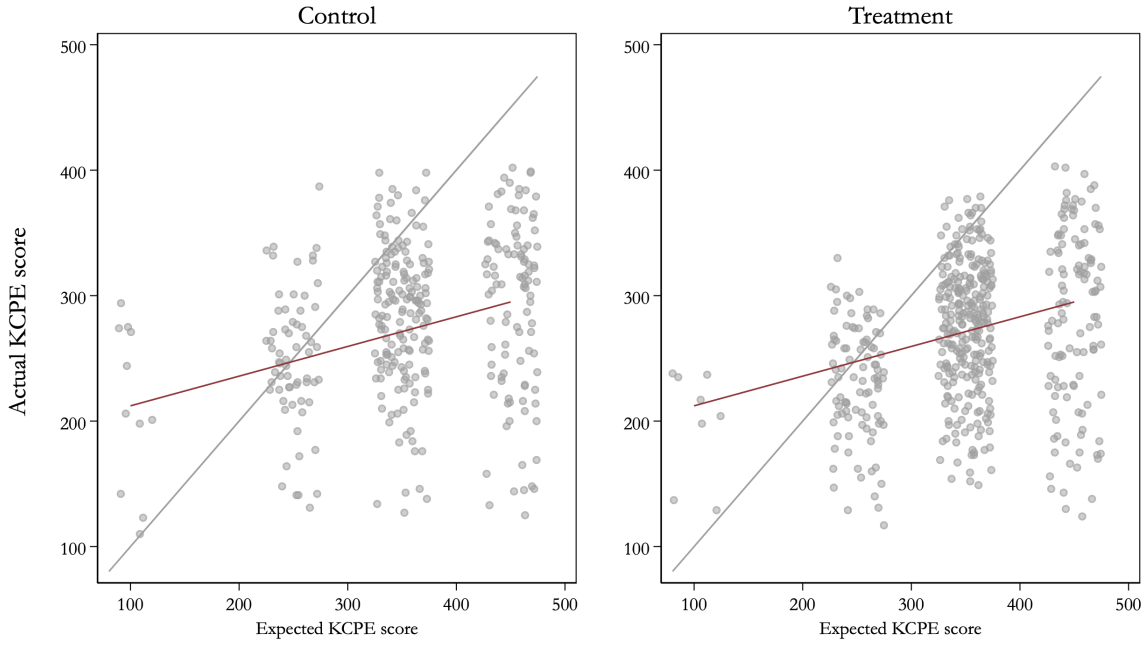


Figure 4: The relationship between actual and expected KCPE performance in Kenya



Notes: The figure presents the relationship between actual and expected KCPE performance in Kenya. Expected KCPE scores were reported in 100-point bins. The figure assigns uniformly distributed random noise in the expectations around the midpoint of the bin. The gray line represents the relationship that would prevail if a one-unit increase in expected test performance were associated with, on average, a one-unit increase in actual test scores. The red line represents the best linear fit to the data.

Table 1: Student-level descriptives and balance

	Uganda				Kenya			
	Control mean	Goal-setting course	Number of students	Number of schools	Control mean	Goal-setting course	Number of students	Number of schools
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Female	0.561	-0.047	327	39	0.504	-0.029	1,963	110
		(0.061)				(0.022)		
		[-0.110]				[-0.014]		
Age at enrollment					11.075	-0.100	1,954	110
						(0.076)		
						[-0.068]		
Years enrolled	1.936	-0.134	327	39	1.721	0.248	1,962	110
		(0.136)				(0.169)		
		[-0.117]				[0.084]		
English score	0.019	-0.048	279	39	0.011	0.037	1,793	110
		(0.182)				(0.090)		
		[-0.046]				[0.148]		
Science score	0.020	0.090	279	39	0.010	-0.045	1,814	110
		(0.248)				(0.088)		
		[0.106]				[0.026]		
Social science score	0.007	0.211	280	39	0.005	-0.103	1,807	110
		(0.233)				(0.100)		
		[0.205]				[-0.059]		
Math score	0.006	-0.041	280	39	0.001	0.039	1,793	110
		(0.281)				(0.094)		
		[-0.086]				[0.094]		
Kiswahili score					0.009	-0.053	1,818	110
						(0.095)		
						[0.029]		
Average score	0.020	0.045	284	39	0.008	-0.035	1,828	110
		(0.231)				(0.087)		
		[0.039]				[0.048]		
F-stat of joint test		0.373				1.439		
p-value		0.728				0.202		

Notes: The table presents descriptive statistics and balance tests for student-level baseline variables for each experiment (Columns 1-4: Uganda and Column 5-8: Kenya). Columns 1 and 5 show the mean of the variable in the control arm. Columns 2 and 6 show the difference in means between the treatment and control groups. Test scores in each subject – Science, Social Science, Math, Kiswahili, English – and experiment are standardized using the respective control group mean. Clustered standard errors are reported in parentheses. Errors are clustered at the strata-level in Kenya due to small randomization strata size following De Chaisemartin and Ramirez-Cuellar (2024) and school-level in Uganda. Normalized differences are reported in square brackets, calculated as the difference between the sample means of experimental arms divided by the square root of the sum of the sample variances. The last rows report the F-stat and p-value of a test that the characteristics are jointly unrelated to the treatment assignment. Stars indicate: *** 1 percent ** 5 percent * 10 percent level of significance. "Age at enrollment" and "Kiswahili" test scores were only available for Kenya sample.

Table 2: Follow-up rate for exams and secondary school transition by experimental arms

	Control mean (1)	Goal-setting course (2)	Number of students (3)	Number of schools (4)
Panel A: Uganda				
Term 2 exam	0.830	0.025	327	39
		(0.065)		
Term 3 exam	0.871	-0.138	327	39
		(0.105)		
National exam (PLE)	0.912	-0.037	327	39
		(0.046)		
Panel B: Kenya				
Term 2 exam	0.813	-0.031	1,963	110
		(0.027)		
Term 3 exam	0.668	0.045	1,963	110
		(0.038)		
National exam (KCPE)	0.820	0.005	1,963	110
		(0.022)		
Secondary school outcomes	0.767	0.020	1,963	110
		(0.031)		
Soft-skill outcomes	0.420	0.055	1,963	110
		(0.066)		

Notes: The table presents follow-up rates for Uganda (Panel A) and Kenya (Panel B). Column 1 shows the follow-up rate in the control group for each outcome - Term 2 Exam, Term 3 Exam, National Exam, secondary school outcomes, self-reported goals, and soft skills (Kenya only). The outcome is an indicator of data availability for the student. Column 2 shows the difference in follow-up rate between the treatment and control arm. Standard errors, clustered at the strata-level in Kenya and school-level in Uganda, are reported in parentheses. Column 3 presents the number of students in the sample – 327 in Uganda and 1963 in Kenya, and Column 4 the number of schools in each experiment – 39 in Uganda and 110 in Kenya. Stars indicate: *** 1 percent ** 5 percent * 10 percent level of significance.

Table 3: Control-group correlations between hard and soft skills
Kenya

	Teacher-graded term 3 score (1)	National exam score (2)	Secondary school enrollment (3)
Baseline score	0.67*	0.68*	0.23*
Grit	0.33*	0.41*	0.11*
Planning	0.24*	0.28*	0.06
Patience	0.16*	0.15*	-0.04
Mental experiencing	0.32*	0.32*	0.14*
Intrinsic motivation	0.05	-0.04	-0.17*
Extrinsic motivation			
External regulation	-0.23*	-0.26*	-0.00
Introjected regulation	-0.20*	-0.21*	-0.10
Identified regulation	0.23*	0.30*	0.14*
Usual daily homework time	0.05	0.02	0.04

Notes: The table reports the pair-wise correlation between characteristics listed on the rows and those listed in column headings. Sample is restricted to the comparison group. Responses were received from 911 students, although sample size varies slightly due to incomplete surveys. * indicates significance at the 5 percent level.

Table 4: Effect of goal-setting course on soft-skills
Kenya

	Grit (1)	Planning (2)	Patience (3)	Mental Experiencing (4)	Joint test p-value (5)
Panel A: Aggregate effect					
Goal-setting course (GSC)	0.346***	0.053	0.160	0.026	
	(0.116)	(0.114)	(0.109)	(0.107)	0.110
Panel B: Interaction with gender					
Goal-setting course (GSC)	0.242*	-0.138	0.006	-0.024	
	(0.127)	(0.145)	(0.128)	(0.122)	0.620
GSC × Female	0.201*	0.363***	0.295***	0.098	
	(0.117)	(0.144)	(0.115)	(0.126)	0.029
Female	-0.160	-0.127	-0.165*	-0.071	
	(0.098)	(0.107)	(0.087)	(0.094)	0.641
Coef: GSC + GSC × Female	0.444	0.226	0.301	0.074	
p-val: GSC + GSC × Female	0.001	0.070	0.011	0.561	0.002
Number of students	911	911	911	904	
Number of schools	84	84	84	84	

Notes: The table presents estimates of the effect of the goal-setting course on soft skills. Panel A reports estimates from Equation 1. Panel B reports estimates interacting assignment to the course with an indicator for whether the student is female. The total effect on female students is reported at the bottom of the table. Data on these outcomes is only available for the Kenya study. All specifications include randomization strata fixed effects. Standard errors are clustered at the stratum level to account for within stratum correlation in assignment to the course (De Chaisemartin and Ramirez-Cuellar, 2024). Column 5 reports p-values from joint tests that all of the coefficients in the row are equal to zero. ***, **, and * indicate significance at 1%, 5%, and 10%.

Table 5: Effect of goal-setting course on motivation and subjective well-being
Kenya

		Extrinsic motivation				
	Intrinsic motivation (1)	External regulation (2)	Introjected regulation (3)	Identified regulation (4)	Subjective well being (5)	Joint test p-value (6)
Panel A: Aggregate effect						
Goal-setting course (GSC)	0.001	-0.031	-0.185**	0.123	-0.001	
	(0.107)	(0.149)	(0.090)	(0.139)	(0.094)	0.715
Panel B: Interaction with gender						
Goal-setting course (GSC)	0.045	0.018	-0.238**	0.086	-0.105	
	(0.127)	(0.170)	(0.116)	(0.157)	(0.103)	0.446
GSC × Female	-0.087	-0.094	0.101	0.073	0.201***	
	(0.114)	(0.113)	(0.112)	(0.105)	(0.077)	0.104
Female	0.134	0.045	-0.011	-0.117	-0.114*	
	(0.100)	(0.096)	(0.093)	(0.088)	(0.059)	0.328
Coef: GSC + GSC × Female	-0.042	-0.076	-0.137	0.158	0.096	
p-val: GSC + GSC × Female	0.722	0.614	0.158	0.260	0.340	0.798
Number of students	909	909	909	909	911	
Number of schools	84	84	84	84	84	

Notes: The table presents estimates of the effect of the goal-setting course on motivations and subjective well being. Panel A reports estimates from Equation 1. Panel B reports estimates interacting assignment to the course with an indicator for whether the student is female. The total effect on female students is reported at the bottom of the table. Data on these outcomes is only available for the Kenya study. All specifications include randomization strata fixed effects. Standard errors are clustered at the stratum level to account for within stratum correlation in assignment to the course (De Chaisemartin and Ramirez-Cuellar, 2024). Column 6 reports p-values from joint tests that all of the coefficients in the row are equal to zero. ***, **, and * indicate significance at 1%, 5%, and 10%.

Table 6: The effect of the goal-setting course on test scores and secondary school enrollment

	Kenya			Uganda	
	Teacher-graded Term 3 score (1)	National Exam (KCPE) (2)	Secondary school enrollment (3)	Teacher-graded Term 3 score (4)	National Exam (PLE) (5)
Panel A: Aggregate effect					
Goal-setting course (GSC)	-0.077	0.011	0.028	-0.072	0.129
	(0.086)	(0.073)	(0.024)	(0.176)	(0.163)
Panel B: Interaction with gender					
Goal-setting course (GSC)	-0.082	0.006	0.033	-0.170	-0.093
	(0.097)	(0.080)	(0.033)	(0.200)	(0.184)
GSC × Female	0.007	0.012	-0.007	0.173	0.409***
	(0.076)	(0.077)	(0.032)	(0.148)	(0.144)
Female	-0.030	0.011	0.040	-0.073	-0.228***
	(0.062)	(0.064)	(0.026)	(0.115)	(0.091)
Control mean	0.000	-0.000	0.873	-0.000	-0.000
Coef: GSC + GSC × Female	-0.075	0.018	0.025	0.003	0.317
p-val: GSC + GSC × Female	0.406	0.830	0.283	0.988	0.067
Observations	6,486	6,870	1,523	1,024	1,172
Number of students	1,335	1,374	1,523	263	293
Number of schools	106	107	110	37	39

Notes: The table presents estimates of the effect of the goal-setting course on test scores and academic advancement. Aggregate effects on test scores in Panel A come from estimating Equation 2, stacking multiple subject-level scores for each student. Estimates of the effect on secondary school enrollment come from estimating Equation 1. Panel B reports analogous specifications, interacting assignment to the course with an indicator for whether the student is female. All specifications include randomization strata fixed effects. In Kenya, standard errors are clustered at the stratum level to account for within-stratum correlation in assignment to the course (De Chaisemartin and Ramirez-Cuellar, 2024). In Uganda, standard errors are clustered at the school level. ***, **, and * indicate significance at 1%, 5%, and 10%.

Table 7: Heterogeneous effects of the goal-setting course on academic achievement by baseline test scores

	Kenya						Uganda			
	Teacher-graded Term 3 score		National exam (KCPE)		Secondary school enrollment		Teacher-graded Term 3 score		National exam (PLE)	
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)
Goal-setting course (GSC)	-0.062	-0.077	0.016	-0.004	0.058	0.060	-0.100	-0.003	0.095	0.242
	(0.087)	(0.114)	(0.068)	(0.061)	(0.047)	(0.040)	(0.190)	(0.223)	(0.166)	(0.186)
GSC × baseline score	-0.006		-0.008		-0.027		-0.153		-0.157*	
	(0.070)		(0.052)		(0.029)		(0.135)		(0.084)	
Baseline score	0.612***		0.567***		0.069***		0.575***		0.589***	
	(0.056)		(0.042)		(0.023)		(0.080)		(0.054)	
GSC × above average score		0.024		0.039		-0.054		-0.181		-0.293
		(0.123)		(0.081)		(0.043)		(0.237)		(0.191)
Above average score		0.968***		0.904***		0.077***		0.901***		0.927***
		(0.101)		(0.071)		(0.028)		(0.164)		(0.145)
Control mean	-0.015	-0.015	-0.003	-0.003	0.872	0.872	0.020	0.020	0.033	0.033
Coef: GSC + GSC × baseline	-0.068		0.008		0.031		-0.253		-0.062	
p-val: GSC + GSC × baseline	0.464		0.936		0.180		0.282		0.724	
Coef: GSC + GSC × above average		-0.053		0.035		0.007		-0.185		-0.051
p-val: GSC + GSC × above average		0.615		0.701		0.773		0.379		0.783
Observations	6,067	6,067	6,406	6,406	1,423	1,405	878	878	1,025	1,025
Number of students	1,258	1,258	1,294	1,294	1,423	1,405	229	229	260	260
Number of schools	106	106	107	107	110	110	35	35	37	37

Notes: The table presents estimates of the effect of the goal-setting course on test scores and academic advancement interacted with measures of initial ability. Odd columns show results that interact assignment to the goal-setting course with the standardized test score. Test scores are centered on the 25th percentile so that the main effect reflects the impact on students between the first and second quartiles. Even columns show analogous results using an interaction with an indicator for whether the student scored above the average of the comparison group. Analysis of exam outcomes is performed at the subject level. For each student-subject observation, the baseline covariate comes from corresponding subject test before assignment. For secondary school enrollment, the the interactions are formed using a composite of the average of scores on all academic subjects. ***, **, and * indicate significance at 1%, 5%, and 10%.

Table 8: Effect of the soft skills course on gendered beliefs about skills
Kenya

	Boys are better at math (1)	Girls are better at math (2)	Boys are better at English (3)	Girls are better at English (4)	More important for boys to attend university (5)	More important for girls to attend university (6)
Panel A: Aggregate effect						
Goal-setting course (GSC)	-0.102***	0.078*	0.037	-0.059	-0.045	-0.032
	(0.039)	(0.047)	(0.030)	(0.042)	(0.037)	(0.026)
Panel B: Interaction with gender						
Goal-setting course (GSC)	-0.141**	0.008	-0.001	-0.045	-0.060	-0.049
	(0.064)	(0.039)	(0.046)	(0.045)	(0.051)	(0.035)
GSC × Female	0.083	0.129***	0.077*	-0.035	0.030	0.031
	(0.075)	(0.040)	(0.046)	(0.076)	(0.048)	(0.033)
Female	-0.197***	0.057**	-0.146***	0.200***	-0.080**	0.009
	(0.056)	(0.027)	(0.031)	(0.064)	(0.038)	(0.027)
Control mean	0.179	0.116	0.120	0.239	0.092	0.054
Coef: GSC + GSC × Female	-0.059	0.137	0.076	-0.079	-0.030	-0.018
p-val: GSC + GSC × Female	0.139	0.019	0.011	0.207	0.433	0.482
Number of students	894	894	894	894	884	884
Number of schools	84	84	84	84	84	84

Notes: This table reports impacts on beliefs about differences in the abilities of girls and boys and gender norms. Each column reports impacts on a different outcome. The outcome in the first two columns are transformations of a response to a question in which students were asked to report which statement they agreed with more: a) girls are better than boys at math, b) girls and boys are equally good in math, c) boys are better than girls in math. The second column reports impacts on outcomes derived from an analogous question regarding English. The last two columns are derived from a question that asked students to indicate which statement they agreed with the most: a) it is more important for a girl to attend university than for a boy; it is equally important for both a girl and a boy to attend university; and it is more important for a boy to attend university than for a girl. ***, **, and * indicate significance at 1%, 5%, and 10%.

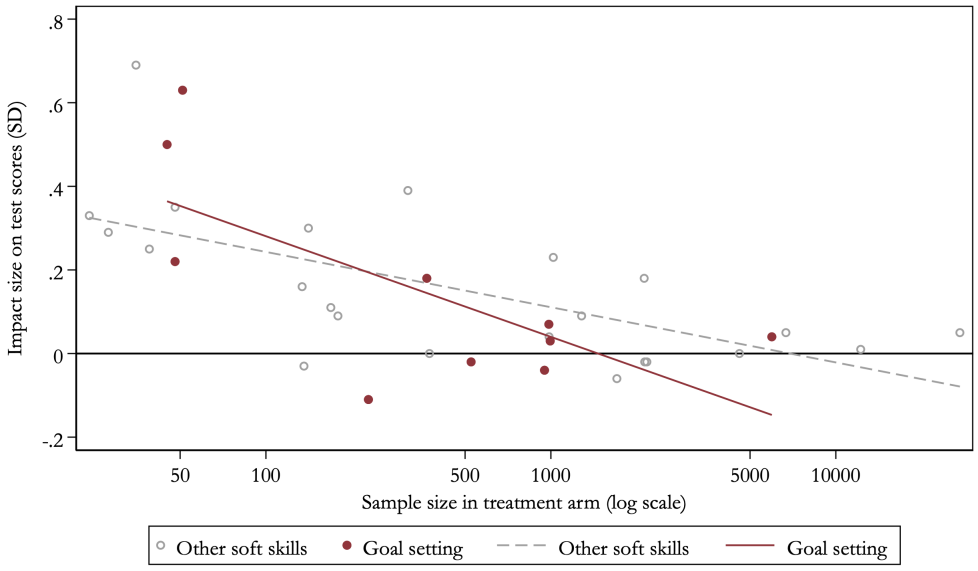
Table 9: Effect of goal-setting course on soft skill surrogate indices
Kenya

	Teacher-graded Term 3 score (1)	KCPE (2)	Secondary school enrollment (3)
Panel A: Aggregate effect			
Goal-setting course (GSC)	0.050*	0.078*	0.011
	(0.030)	(0.047)	(0.021)
Equal to effect on outcome (p-value)	0.220	0.498	0.552
Panel B: Interaction with gender			
Goal-setting course (GSC)	0.017	0.037	0.010
	(0.031)	(0.047)	(0.022)
GSC × Female	0.065	0.079	0.001
	(0.031)	(0.047)	(0.022)
Female	-0.030	-0.032	-0.003
	(0.031)	(0.047)	(0.022)
Coef: GSC + GSC × Female	0.082	0.116	0.011
p-val: GSC + GSC × Female	0.014	0.020	0.657
Equal to effect on outcome (p-value):			
GSC	0.458	0.766	0.660
GSC× female	0.186	0.440	0.887
Female	0.656	0.693	0.373
GSC + GSC × Female	0.167	0.460	0.768
Number of scores	21,015	21,015	847
Number of students	848	848	847
Number of schools	84	84	83

Notes: The table presents effect of goal-setting course on soft skill surrogate indices on test scores in Kenya. Panel A reports estimates from Equation 1. Panel B reports estimates interacting assignment to the course with an indicator for whether the student is female. The total effect on female students is reported at the bottom of the table. Data on these outcomes is only available for the Kenya study. All specifications include randomization strata fixed effects. Standard errors are clustered at the stratum level. ***, **, and * indicate significance at 1%, 5%, and 10%.

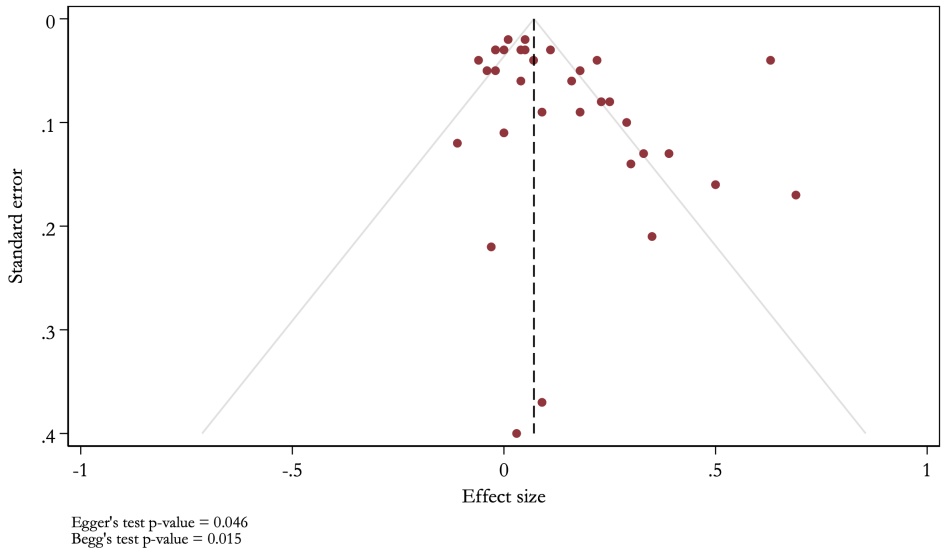
A Online Appendix

Figure A1: Efficacy versus scale for goal-setting interventions



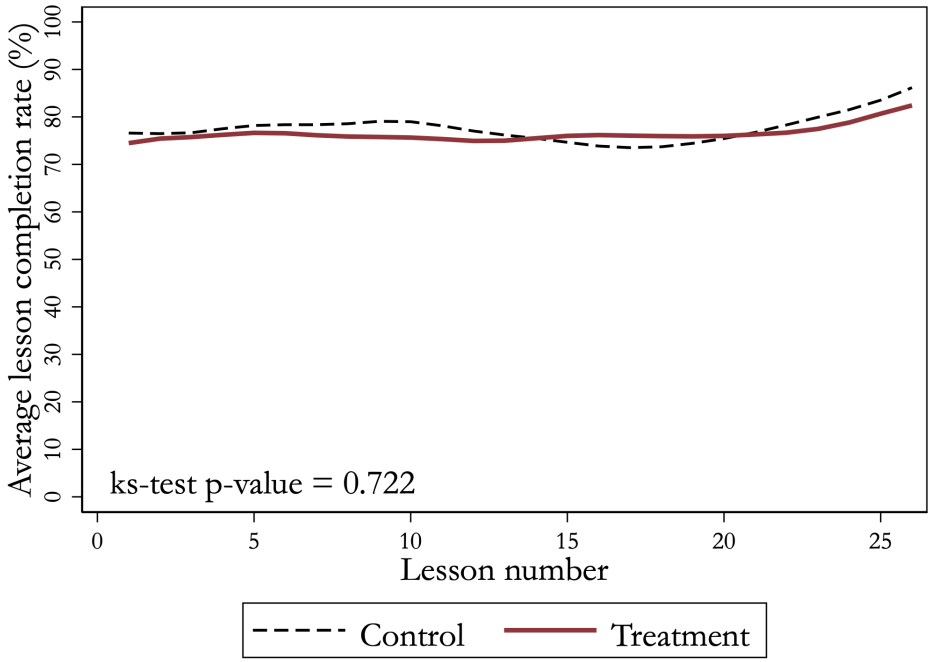
Notes: The figure shows the relationship between the effect on test scores and their scale,i.e., number of students in treatment group, for goal-setting (red) and other soft-skill interventions (grey) separately. Additional details are in Table A1.

Figure A2: Funnel plot of randomized evaluations of soft-skills education



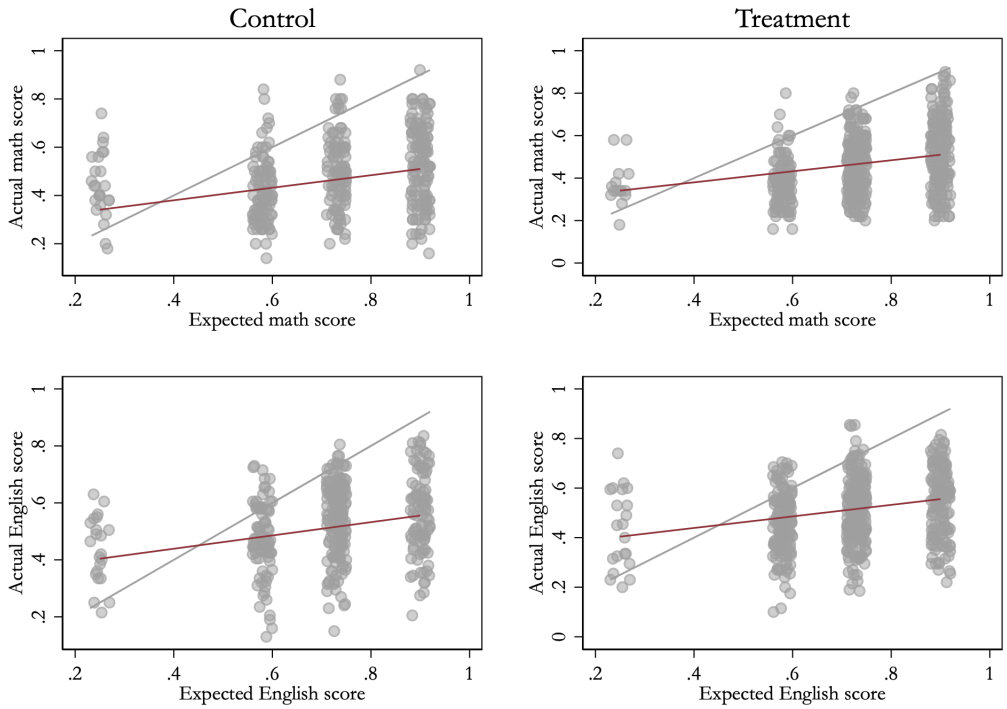
Notes: The figure presents funnel plots of the 33 randomized evaluations of soft-skills education included in Figure 1. P-values for Egger and Begg test are presnted at the bottom of the graph. Additional details are in Table A1.

Figure A3: Average lesson completion rates during the curriculum over time by experimental arms



Notes: The figure presents density plots of average lesson completion rate per lesson over time, from the administrative data in Kenya. Lesson completion for each lesson of the 25 goal-setting lessons in treatment schools (red) or 25 corresponding revision lessons in control schools (gray), are calculated by averaging over schools for each lesson number in each experimental arm. Average lesson completion is shown on the y-axis and the lesson number is on the x-axis. In the bottom left corner of the graph, the p-value from a Kolmogorov–Smirnov test for differences in distributions between experimental arms is presented.

Figure A4: The relationship between actual and expected performance on teacher-administered end-of-term exams in Kenya



Notes: The figure presents the relationship between actual and expectedend-of-year performance in Kenya. The red line represents the best linear fit to the data. Expected end-of-year performance on teacher-graded math and English scores were reported in bins (less than 50 percent, 51-65, 66-80, and more than 80 percent). The figure assigns uniformly distributed random noise in the expectations around the midpoint of the bin. The gray line represents the relationship that would prevail if a one-unit increase in expected test performance were associated with, on average, a one-unit increase in actual test scores. The red line represents the best linear fit to the data.

Table A1: Overview of literature on soft-skills interventions in education

Citation	Treatment Arm (N)	Total (N)	Unit of randomization	Impact Size (SD)	Duration (hours)	Type
Goal-Setting Interventions						
Chase et al. (2013)	48	132	individual	0.22	0.75	Goal Setting
				(0.040)		
Chase et al. (2013)	51	132	individual	0.63***	0.75	Goals Setting & Value Affirmation
				(0.040)		
Clark et al. (2020)† — Study 1	995	1967	individual	0.028	0.50	Goal Setting (Performance Goals)
				(0.398)		
Clark et al. (2020)† — Study 2	984	2004	individual	0.068*	0.50	Goal Setting (Task Goals)
				(0.043)		
Dobronyi et al. (2019)	229	1492	individual	-0.111	2.00	Goal Setting
				(0.124)		
Islam et al. (2024)‡	5962	13067	school	0.042	0.50	Goal Setting
				(0.062)		
Morisano et al. (2010)†	45	85	individual	0.50**	2.50	Goal Setting
				(0.163)		
Oreopoulos and Petronijevic (2019) — Study 2	950	1591	individual	-0.043	2.00	Goal Setting
				(0.054)		
van Lent (2019)	525	2100	individual	-0.02	0.10	Goal Setting
				(0.050)		
van Lent and Souverijn (2020)	367	756	individual	0.181**	1.50	Goal Setting
				(0.091)		
Other Soft-Skills Interventions						
Alan et al. (2019)	2126	3200	school	0.183	24.00	Grit Intervention
				(0.050)		
Aronson et al. (2002)†	28	79	individual	0.29***	3.00	Growth Mindset
				(0.096)		
Berger et al. (2025)	134	572	classroom	0.161	12.50	Working Memory
				(0.056)		
Bettinger et al. (2018)	179	354	individual	0.089	2.25	Growth Mindset
				(0.086)		
Blackwell et al. (2007)	48	91	individual	0.35*	3.25	Growth Mindset
				(0.212)		
Broda et al. (2018)† — Study 1	2172	6529	individual	-0.02	0.25	Social Belonging
				(0.030)		
Broda et al. (2018)† — Study 2	2135	6529	individual	-0.02	0.25	Growth Mindset
				(0.030)		
Cook et al. (2014)†	24	58	individual	0.332**	27.00	Social Cognitive Skills
				(0.134)		
Dee (2015)	1282	2564	individual	0.09	0.25	Self Affirmation
				(0.365)		
Duckworth et al. (2007)	39	77	individual	0.25**	3.00	Implementation Intention
				(0.083)		
Ganimian (2020)	4590	8865	school	0.002	1.00	Growth Mindset
				(0.032)		
Good et al. (2003)†	35	138	school	0.685**	3.00	Growth Mindset
				(0.174)		
Huillery et al. (2025)	12228	24142	cohort	0.01	11.00	Growth Mindset
				(0.020)		
Hulleman and Harackiewicz (2009)†	136	262	individual	-0.03	2.00	Relevance
				(0.220)		
Humphrey et al. (2015)	1705	3336	school	-0.061	20.00	PATHS
				(0.036)		
Oreopoulos and Petronijevic (2019) — Study 3	986	2437	individual	0.039	2.00	Growth Mindset
				(0.026)		
Outes-Leon et al. (2020)	27255	54510	school	0.047*	1.50	Growth Mindset
				(0.020)		
Oyserman et al. (2006)†	141	264	individual	0.30**	13.00	Possible Selves
				(0.139)		
Porter et al. (2022)	1020	1996	teacher	0.23***	12.50	Growth Mindset
				(0.082)		
Schunk et al. (2022)	315	572	classroom	0.391***	4.00	Self-regulation training
				(0.132)		
Sorrenti et al. (2024)	375	1675	school	-0.004	45.00	PATHS
				(0.110)		
Yeager et al. (2014)†	169	338	individual	0.11***	0.50	Self-Transcendent Purpose
				(0.034)		
Yeager et al. (2019)†	6680	13360	individual	0.052***	1.00	Growth Mindset
				(0.030)		

Notes: The table presents details from 33 randomized evaluation of soft skill interventions in education at the treatment arm level. Fifteen studies reported a single test score outcome. For studies with multiple outcomes, we applied the following decision rules to obtain a single point estimate: (i) if the intervention focused on a specific subject or exam, we used that assessment (1 study); (ii) where multiple outcomes qualified, we prioritized research-designed tests, followed by standardized exams (e.g., national exam), then internal measures such as GPA or grades (3 studies); (iii) if assessments were administered at multiple points in time, we selected the measure closest to one year post-intervention (11 studies); and (iv) where multiple subject outcomes remained after the above criteria were applied, we constructed a precision-weighted average (8 studies). The table reports treatment-arm sample size, total study sample size, intervention duration (in hours), and unit of randomization. Interventions are grouped into common categories from the education literature. Test score impacts are expressed in standard deviations, with standard errors in parentheses. Stars indicate: *** 1 percent ** 5 percent * 10 percent level of significance.
† Indicates that we did not receive a confirmation from the original authors after attempted contact
‡ The interventions were conducted with students outside of formal school settings

Table A2: 2022 national exam test scores at the national level and in Bridge schools

National exam score	Bridge (1)	National (2)
400-500	0.13%	0.77%
300-399	31.91%	24.94%
200-299	48.60%	50.22%
100-199	19.36%	24.02%
001-099	0%	0.06%
Registered for exam	749	1,244,335

Notes: The table presents percentage of students at various bandwidths of national exam (KCPE) test scores in Bridge schools as compared to the national performance in 2022. Percentages for students in Bridge schools are calculated using control group data. National percentages come from the Ministry of Education.

Table A3: Student-level descriptives and balance
Conditional on taking national exam

	Uganda				Kenya			
	Control mean	Goal-setting course	Number of students	Number of schools	Control mean	Goal-setting course	Number of students	Number of schools
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Female	0.558	-0.016	293	39	0.505	-0.035*	1,611	110
		(0.058)				(0.022)		
		[-0.064]				[-0.019]		
Age at enrollment					11.073	-0.091	1,603	110
						(0.081)		
						[-0.084]		
Years enrolled	2.051	-0.167	293	39	1.822	0.227	1,611	110
		(0.130)				(0.178)		
		[-0.160]				[0.076]		
English score	0.029	-0.037	256	39	0.044	0.020	1,481	110
		(0.200)				(0.091)		
		[-0.037]				[0.155]		
Science score	0.002	0.115	256	39	0.051	-0.059	1,493	110
		(0.247)				(0.095)		
		[0.120]				[0.034]		
Social science score	0.001	0.306	257	39	0.045	-0.127	1,489	110
		(0.234)				(0.105)		
		[0.275]				[-0.070]		
Math score	-0.002	-0.021	256	39	0.048	0.040	1,482	110
		(0.274)				(0.102)		
		[-0.098]				[0.100]		
Kiswahili score					0.054	-0.081	1,497	110
						(0.099)		
						[0.020]		
Average score	0.013	0.101	260	39	0.058	-0.056	1,504	110
		(0.226)				(0.091)		
		[0.073]				[0.046]		
F-stat of joint test		0.833				1.439		
p-value		0.659				0.202		

Notes: The table presents descriptive statistics and balance tests on student-level baseline variables for each experiment (Columns 1-4: Uganda and Column 5-8: Kenya), for the sample of students we were able to follow-up in the national exam (Columns 1-4: Uganda and Column 5-8: Kenya). Columns 1 and 5 show the mean of the variables in the control groups. Columns 2 and 6 show the difference in means between the treatment and control arms. Test scores in each subject – science, Social studies, math, Kiswahili, English – and experiment are standardized using the respective control group mean. Clustered standard errors are reported in parentheses. Errors are clustered at the strata-level in Kenya due to small randomization strata size following De Chaisemartin and Ramirez-Cuellar (2024) and school-level in Uganda. Normalized differences are reported in square brackets, calculated as the difference between the sample means of experimental arms divided by the square root of the sum of the sample variances. The last rows report the F-stat and p-value of a test that the characteristics are jointly unrelated to the treatment assignment. Stars indicate: *** 1 percent ** 5 percent * 10 percent level of significance. "Age at enrollment" and "Kiswahili" test scores were only available for Kenya sample.

Table A4: Student-level descriptives and balance
Conditional on completing soft skills survey

	Control mean	Goal-Setting course	Number of students	Number of schools
	(1)	(2)	(3)	(4)
Female	0.525	-0.026	911	85
		(0.034)		
		[0.007]		
Age at enrollment	11.067	-0.097	905	85
		(0.133)		
		[-0.086]		
Years enrolled	2.208	-0.058	911	85
		(0.318)		
		[-0.151]		
English score	0.133	-0.077	845	85
		(0.133)		
		[-0.005]		
Science score	0.014	0.040	850	85
		(0.113)		
		[0.002]		
Social science score	-0.003	0.034	850	85
		(0.116)		
		[-0.080]		
Math score	0.117	0.189	845	85
		(0.137)		
		[0.066]		
Average score	0.074	0.006	856	85
		(0.126)		
		[-0.021]		
Kiswahili score	0.066	-0.259*	853	85
		(0.141)		
		[-0.080]		
F-stat of joint test		1.726		
p-value		0.114		

Notes: The table presents descriptive statistics and balance tests for students who have completed the soft skills survey in Kenya. Columns 1 show the mean of the variable in the control arm. Columns 2show the difference in means between the treatment and control groups. Test scores in each subject – science, social studies, math, Kiswahili, English – and experiment are standardized using the respective control group mean. Clustered standard errors are reported in parentheses. Errors are clustered at the strata-level in Kenya due to small randomization strata size following De Chaisemartin and Ramirez-Cuellar (2024) and school-level in Uganda. Normalized differences are reported in square brackets, calculated as the difference between the sample means of experimental arms divided by the square root of the sum of the sample variances. The last rows report the F-stat and p-value of a test that the characteristics are jointly unrelated to the treatment assignment. Stars indicate: *** 1 percent ** 5 percent * 10 percent level of significance. "Age at enrollment" and "Kiswahili" test scores were only available for Kenya sample.

Table A5: Actual versus expected secondary school type

	Expected secondary school type				
Actual enrollment	National (1)	Extra-county (2)	County (3)	Sub-county (4)	Row total (5)
National	0.046	0.006	0.000	0.000	0.052
Extra-county	0.224	0.033	0.013	0.002	0.273
County	0.167	0.026	0.014	0.000	0.208
Sub-county	0.256	0.043	0.026	0.015	0.340
Private	0.025	0.004	0.002	0.000	0.031
Not enrolled	0.063	0.020	0.008	0.005	0.096
Column total	0.782	0.132	0.064	0.023	1.000

Notes: This table shows the joint distribution between expected and actual secondary school enrollment type. Each row-column cell reports the fraction of the total sample that expected to enroll in a particular type (columns) and the actual type of secondary school they enrolled in (rows). The last row shows the total share that expected to go to each type and the last column shows the total share that actually enrolled in each type.

Table A6: Control-group correlations between expectations and soft skills
Kenya

	Expected KCPE score (1)	Actual minus expected KCPE (2)	Distance from expectation (3)	Underestimated performance (4)	Overestimated performance (5)	Overestimated performance by 100 points (6)
Baseline score	0.23*	0.30*	-0.31*	0.08	-0.08	-0.30*
Soft-skills						
Grit	0.24*	0.08	-0.21*	-0.01	0.01	-0.07
Planning	0.19*	-0.01	-0.16*	-0.13*	0.13*	-0.04
Patience	0.09	0.01	-0.03	-0.01	0.01	-0.08
Mental experiencing	0.21*	0.04	-0.08	-0.06	0.06	-0.05
Intrinsic motivation	-0.09	0.05	0.05	0.07	-0.07	-0.00
Extrinsic motivation						
External regulation	-0.10	-0.06	0.07	-0.05	0.05	0.13*
Introjected regulation	-0.11*	-0.05	0.13*	0.01	-0.01	0.05
Identified regulation	0.17*	0.04	-0.14*	-0.00	0.00	-0.10
Usual weekly homework time	-0.04	0.06	0.02	0.06	-0.06	-0.02

Notes: The table reports the pair-wise correlation between characteristics listed on the rows and those listed in column headings. Sample is restricted to the comparison group. Responses were received from 911 students, although sample size varies slightly due to incomplete surveys. * indicates significance at the 5 percent level.

Table A7: Effect of the goal-setting course on national exam expectations
Kenya

	Expected KCPE score (1)	Actual minus expected KCPE (2)	Distance from expectation (3)	Underestimated performance (4)	Overestimated performance (5)	Overestimated performance by 100 points (6)
Panel A: Aggregate effect						
Goal-setting course (SSC)	-6.186	3.759	-11.463	-0.002	0.002	-0.096
	(8.961)	(9.529)	(10.363)	(0.041)	(0.041)	(0.067)
Panel B: Interaction with gender						
Goal-setting course (SSC)	-4.514	2.686	-8.414	0.006	-0.006	-0.069
	(10.099)	(9.155)	(9.026)	(0.048)	(0.048)	(0.064)
SSC × Female	-3.272	2.382	-6.398	-0.018	0.018	-0.055
	(9.132)	(11.817)	(9.156)	(0.052)	(0.052)	(0.071)
Female	5.295	-5.956	10.979*	0.024	-0.024	0.052
	(7.264)	(9.054)	(6.438)	(0.044)	(0.044)	(0.052)
Control mean	351.156	-80.079	91.824	0.142	0.858	0.372
Coef: SSC + SSC × Female	-7.786	5.068	-14.812	-0.011	0.011	-0.124
p-val: SSC + SSC × Female	0.441	0.696	0.263	0.819	0.819	0.151
Number of students	907	830	830	830	830	830
Number of schools	84	84	84	84	84	84

Notes: The table presents estimates of the effect of the goal-setting course on expectations of test scores on the national KCPE exam. Panel A reports estimates from Equation 1. Panel B reports estimates interacting assignment to the course with an indicator for whether the student is female. The total effect on female students is reported at the bottom of the table. Expected KCPE scores were reported in 100-point bins. The first column is a continuous mid-point of the expected KCPE score. The second column is the difference between actual KCPE score and the midpoint of the expected KCPE score. The third column is the absolute value of this difference. Columns 4-6 are are binary variables taking on value one if the student underestimates, overestimates or overestimates by more than 100 points their expected score relative to their actual score. Data on these outcomes is only available for the Kenya study. All specifications include randomization strata fixed effects. Standard errors are clustered at the stratum level to account for within stratum correlation in assignment to the course (De Chaisemartin and Ramirez-Cuellar, 2024). ***, **, and * indicate significance at 1%, 5%, and 10%.

Table A8: Effect of the soft skills course on time spent on homework
Kenya

	Homework time(min) (1)	Homework time(min) (2)	Homework time(min) (3)
Goal-setting course (GSC)	-1.37	3.84	-1.34
	(4.62)	(4.76)	(4.44)
GSC × female		-10.02**	
		(4.94)	
Female	7.15***	13.12***	7.14***
	(2.07)	(3.57)	(2.07)
GSC × baseline score			-0.22
			(3.01)
Baseline score	-2.24	-2.30*	-2.11
	(1.41)	(1.37)	(2.52)
Control mean	62.68	62.68	62.68
Coef: GSC + GSC × female		-6.18	
p-val: GSC + GSC × female		0.28	
Coef: GSC + GSC × baseline			-1.56
p-val: GSC + GSC × baseline			0.80
Number of students	852	852	852
Number of schools	83	83	83

Notes: Table shows the effect of the soft-skills course on the amount of time students report spending on homework on a usual day in minutes. Baseline Test scores are centered on the 25th percentile so that the main effect reflects the impact on students between the first and second quartiles. Stars ***, **, and * indicate significance at 1%, 5%, and 10%.

Table A9: Effect of the goal-setting course on test scores pooled for both experiments

	Teacher-graded end-of-year exam score (1)	National Exam (KCPE) (2)
Goal-setting course (GSC)	-0.084	0.027
	(0.063)	(0.055)
Control Mean	0.029	0.042
Observations	7,510	8,042
Number of students	1,598	1,667
Number of schools	143	146

Notes: The table presents estimates of the effect of the goal-setting course on test scores, pooling the Uganda and Kenya data. Aggregate effects on test scores come from estimating an equation similar to Equation 2 by stacking experiment and subject-level data. All specifications include randomization strata, subject and experiment fixed effects. Standard errors are clustered at the school level. ***, **, and * indicate significance at 1%, 5%, and 10%.

Table A10: The effect of the goal-setting course on additional post-primary outcomes
Kenya

	Received scholarship (1)	National (2)	Extra-county (3)	County (4)	Sub-county (5)	Private sector (6)
Panel A: Aggregate effect						
Goal-setting course (GSC)	-0.007	0.018	-0.012	0.027	0.066	-0.021*
	(0.016)	(0.024)	(0.039)	(0.043)	(0.072)	(0.011)
Panel B: Interaction with gender						
Goal-setting course (GSC)	-0.003	0.031	0.004	0.042	0.080	-0.024***
	(0.019)	(0.028)	(0.045)	(0.052)	(0.083)	(0.010)
GSC × Female	-0.007	-0.028	-0.033	-0.028	-0.023	0.009
	(0.025)	(0.023)	(0.046)	(0.044)	(0.077)	(0.015)
Female	0.017	0.014	0.003	0.049	0.095*	0.010
	(0.018)	(0.015)	(0.038)	(0.034)	(0.054)	(0.013)
Control mean	0.052	0.059	0.349	0.534	0.719	0.026
Coef: GSC + GSC × Female	-0.010	0.004	-0.029	0.014	0.057	-0.015
p-val: GSC + GSC × Female	0.625	0.888	0.524	0.748	0.474	0.362
Number of students	1,355	1,433	1,433	1,433	1,433	1,433
Number of schools	109	110	110	110	110	110

Notes: Table reports impacts on an indicator for whether a student received a scholarship for secondary school and the type of secondary school enrolled. Public secondary schools in Kenya are classified as either national, extra-county, county, or sub-county. National schools tend to be most competitive, followed by extra-county, county, and sub-county. For schools that are less competitive than national, we examine the impact on enrollment in that type *or a more competitive type*. Students enrolled in private school have a value of zero because the ordinaly ranking of private among the public school types is not clear. For all secondary school types, students who did not enroll in any secondary school have a value of zero. ***, **, and * indicate significance at 1%, 5%, and 10%.

Table A11: The effect of the goal-setting course on test scores
Girls only

	Kenya						Uganda			
	Teacher-graded end-of-year exam		National exam (KCPE)		Secondary school enrollment		Teacher-graded end-of-year exam		National exam (PLE)	
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)
Goal-setting course (GSC)	-0.051	-0.083	0.035	-0.034	0.031	0.002	-0.047	0.065	0.238	0.446**
	(0.077)	(0.085)	(0.080)	(0.077)	(0.045)	(0.039)	(0.193)	(0.245)	(0.160)	(0.205)
GSC × baseline score	-0.015		0.027		0.000		-0.156		-0.219**	
	(0.066)		(0.055)		(0.027)		(0.163)		(0.101)	
Baseline score	0.612***		0.532***		0.039*		0.574***		0.580***	
	(0.047)		(0.038)		(0.021)		(0.094)		(0.067)	
GSC × above average score		0.023		0.101		0.040		-0.240		-0.455*
		(0.124)		(0.095)		(0.052)		(0.283)		(0.236)
Above average score		0.956***		0.855***		-0.017		0.909***		0.973***
		(0.100)		(0.065)		(0.029)		(0.196)		(0.187)
Control mean	0.007	0.007	0.065	0.065	0.895	0.894	-0.030	-0.030	-0.068	-0.068
Coef: GSC + GSC × baseline	-0.066		0.062		0.031		-0.203		0.019	
p-val: GSC + GSC × baseline	0.525		0.557		0.264		0.434		0.916	
Coef: GSC + GSC × above average		-0.061		0.067		0.042		-0.175		-0.009
p-val: GSC + GSC × above average		0.621		0.529		0.252		0.438		0.962
Observations	3,027	3,027	3,216	3,216	705	693	511	511	564	564
Number of students	627	627	649	649	705	693	131	131	143	143
Number of schools	101	101	102	102	109	109	32	32	33	33

Notes: The table presents estimates of the effect of the goal-setting course on test scores and academic advancement interacted with measures of initial ability. Odd columns show results that interact assignment to the goal-setting course with the standardized test score. Test scores are centered on the 25th percentile so taht the main effect reflects the impact on students between the first and second quartiles. Even columns show analogous results using an interaction with an indicator for whether the student scored above the average of the comparison group. Analysis of exam outcomes is performed at the subject level. For each student-subject observation, the baseline covariate comes from corresponding subject test before assignment. For secondary school enrollment, the the interactions are formed using a composite of the average of scores on all academic subjects. ***, **, and * indicate significance at 1%, 5%, and 10%.

Table A12: Heterogeneous effects of the goal-setting course on soft skills by baseline test scores

						Extrinsic motivation			
	Grit	Planning	Patience	Mental Experiencing	Intrinsic motivation	External regulation	Introjected regulation	Identified regulation	Subjective wellbeing
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
Goal setting course (GSC)	0.325***	0.018	0.095	0.033	-0.040	-0.013	-0.190*	0.143	-0.018
	(0.126)	(0.111)	(0.104)	(0.113)	(0.106)	(0.150)	(0.099)	(0.144)	(0.086)
GSC × baseline score	-0.036	-0.034	0.201***	-0.079	-0.003	0.149**	0.071	-0.128	0.049
	(0.096)	(0.083)	(0.074)	(0.080)	(0.089)	(0.073)	(0.066)	(0.079)	(0.047)
Baseline score	0.330***	0.186***	0.201***	0.315***	0.001	-0.228***	-0.238***	0.272***	0.140***
	(0.071)	(0.065)	(0.070)	(0.054)	(0.068)	(0.066)	(0.061)	(0.062)	(0.030)
Control mean	-0.000	-0.001	0.003	0.025	-0.004	-0.006	-0.009	0.009	0.001
Coef: GSC + GSC × baseline	0.288	-0.016	0.296	-0.045	-0.043	0.135	-0.119	0.015	0.031
p-val: GSC + GSC × baseline	0.052	0.918	0.048	0.681	0.781	0.395	0.228	0.927	0.787
Observations	856	856	856	849	854	854	854	854	856
Number of students	856	856	856	849	854	854	854	854	856
Number of schools	83	83	83	83	83	83	83	83	83

Notes: The table presents estimates of the effect of the goal-setting course on test scores and academic advancement interacted with measures of initial ability. Odd columns show results that interact assignment to the goal-setting course with the standardized test score. Test scores are centered on the 25th percentile so taht the main effect reflects the impact on students between the first and second quartiles. ***, **, and * indicate significance at 1%, 5%, and 10%.

Table A13: Effect of goal-setting course on national exams

	Math (1)	Science (2)	Social science (3)	English (4)	Kiswahili (5)
Panel A: Uganda					
Aggregate effect:					
Goal-setting course	0.323	0.181	0.002	0.005	
	(0.200)	(0.179)	(0.193)	(0.197)	
By Gender:					
Goal-setting course (GSC)	0.143	-0.099	-0.114	-0.292	
	(0.232)	(0.237)	(0.204)	(0.220)	
GSC × Female	0.292	0.495**	0.324*	0.483***	
	(0.183)	(0.224)	(0.171)	(0.167)	
Female	-0.232*	-0.384***	-0.299**	-0.018	
	(0.120)	(0.116)	(0.138)	(0.080)	
Coef: GSC + GSC × Female	0.435	0.396	0.210	0.191	
p-val: GSC + GSC × Female	0.050	0.058	0.361	0.366	
Number of students	293	293	293	293	
Number of schools	39	39	39	39	
Panel B: Kenya					
Aggregate effect:					
Goal-setting course	-0.103	-0.030	0.053	0.087	0.056
	(0.084)	(0.083)	(0.090)	(0.096)	(0.098)
By Gender:					
Goal-setting course (GSC)	-0.144	-0.017	-0.005	0.125	0.042
	(0.095)	(0.091)	(0.093)	(0.097)	(0.104)
GSC × Female	0.150**	-0.063	0.042	-0.002	-0.011
	(0.071)	(0.077)	(0.080)	(0.079)	(0.096)
Female	-0.160***	-0.168***	-0.129**	0.094	0.172**
	(0.051)	(0.065)	(0.063)	(0.066)	(0.076)
Coef: GSC + GSC × Female	0.005	-0.080	0.038	0.123	0.032
p-val: GSC + GSC × Female	0.953	0.390	0.702	0.240	0.772
Number of students	1,397	1,397	1,397	1,397	1,397
Number of schools	108	108	108	108	108

Notes: The table presents estimates of the effect of the goal-setting course on subject-level test scores on the national KCPE exam for Kenya and PLE exam for Uganda. Panel A presents data from the Uganda sample and Panel B from the Kenya sample. "Aggregate effects" come from estimating Equation 1 for subject-level scores. "By gender" reports analogous specifications, interacting assignment to the course with an indicator for whether the student is female. The total effect on female students is reported at the bottom of the panel. All specifications include randomization strata fixed effects. In Kenya, standard errors are clustered at the stratum level to account for within-stratum correlation in assignment to the course (De Chaisemartin and Ramirez-Cuellar, 2024). In Uganda, standard errors are clustered at the school level. ***, **, and * indicate significance at 1%, 5%, and 10%.

Table A14: Effect of goal-setting course on teacher-graded end-of-year exams

	Math (1)	Science (2)	Social science (3)	English (4)	Kiswahili (5)
Panel A: Uganda					
Aggregate effect:					
Goal-setting course	0.052	-0.354*	-0.048	0.064	
	(0.223)	(0.200)	(0.205)	(0.192)	
By Gender:					
Goal-setting course (GSC)	-0.082	-0.562***	-0.010	-0.181	
	(0.244)	(0.226)	(0.234)	(0.227)	
GSC × Female	0.137	0.300	0.037	0.283	
	(0.185)	(0.187)	(0.187)	(0.185)	
Female	-0.020	-0.200	-0.056	-0.036	
	(0.155)	(0.168)	(0.166)	(0.111)	
Coef: GSC + GSC × Female	0.055	-0.262	0.027	0.102	
p-val: GSC + GSC × Female	0.824	0.213	0.909	0.607	
Number of students	253	259	259	253	
Number of schools	36	37	37	36	
Panel B: Kenya					
Aggregate effect:					
Goal-setting course	-0.123	-0.007	-0.128	-0.105	-0.042
	(0.105)	(0.096)	(0.099)	(0.099)	(0.112)
By Gender:					
Goal-setting course (GSC)	-0.175*	-0.025	-0.167	-0.078	-0.040
	(0.106)	(0.110)	(0.104)	(0.105)	(0.132)
GSC × Female	0.166**	0.028	0.030	0.003	-0.080
	(0.084)	(0.108)	(0.094)	(0.073)	(0.100)
Female	-0.135**	-0.169**	-0.078	0.053	0.110
	(0.067)	(0.085)	(0.069)	(0.058)	(0.081)
Coef: GSC + GSC × Female	-0.009	0.003	-0.137	-0.075	-0.119
p-val: GSC + GSC × Female	0.935	0.977	0.182	0.360	0.256
Number of students	1,305	1,314	1,353	1,309	1,331
Number of schools	108	106	107	108	107

Notes: The table presents estimates of the effect of the goal-setting course on subject-level test scores on the teacher-graded end-of-year exam. Panel A presents data from the Uganda sample and Panel B from the Kenya sample. "Aggregate effects" come from estimating Equation 1 for subject-level scores. "By gender" reports analogous specifications, interacting assignment to the course with an indicator for whether the student is female. The total effect on female students is reported at the bottom of the panel. All specifications include randomization strata fixed effects. In Kenya, standard errors are clustered at the stratum level to account for within-stratum correlation in assignment to the course (De Chaisemartin and Ramirez-Cuellar, 2024). In Uganda, standard errors are clustered at the school level. ***, **, and * indicate significance at 1%, 5%, and 10%.

A.1 Lesson plan of goal-setting course

1. Setting goals
2. Short-term goals
3. Evaluating your goals
4. Procrastination
5. Evaluating your progress
6. Goal adjustment
7. Setting medium-term and long-term goals
8. Visualizing goals
9. Setting important goals
10. Ranking goals
11. Evaluating goals
12. Goal adjustment
13. Goal ranking
14. Goal prioritization
15. How to prioritize personal goals
16. Managing your time
17. Measuring progress
18. Making changes to your goals
19. Scheduling
20. Strategies and goal setting
21. Achieving personal goals
22. Evaluating progress on your goals
23. Making changes to your goals
24. Breaking down your goals
25. Obstacles to goals

A.2 Measures of soft skills, motivation, and well-being

An index with K items is constructed, by taking the average of the control-group standardized scores (z-scores) of each item. That is, we compute Equation 5, where μ_{0k} and σ_{0k} are the estimated control-group mean and standard deviation for item k in index i:

$$\tilde{y}_i = \frac{1}{K} \sum^{K} \left(\frac{y_{ik} - \mu_{0k}}{\sigma_{0k}} \right) \tag{5}$$

Our estimates for these indices represents standard deviation changes relative to the control group.

1. **Patience**: Items i-vi are based on Considerations of Future Consequences Scale (Samek et al., 2021; Strathman et al., 1994) and item vii is based on Falk et al. (2023). Students can choose between four options on a likert-scale – very much like me; a bit like me ; not really like me; not at all like me. Item i-ii and vii are scored as 4,3,2, and 1, respectively, and items iii-vi are reverse scored, i.e., 1-4 respectively. Such that, larger numbers are associated with greater patience. We construct an index by computing Equation 5.

    (a) Patience index:

        i. I first think about what could happen in the future. Then, I try to reach those future objectives by doing specific activities right now.

        ii. I often do things now, in order to achieve objectives that may happen many years from now.

        iii. How good I feel right now is a big factor in the decisions I make or the actions I take.

        iv. I think that giving up something now is usually unnecessary, since I can deal with my future objectives at a later time.

v. I only focus on the problems I have right now, thinking that I will take care of any future problems at a later date.

vi. I only do activities that have results right now, because I think these are more important than activities that have results in the future.

vii. In comparison to others, are you a person who is generally willing to give up something today in order to get a benefit in the future?

2. **Grit**: Students can choose between four options on a likert-scale of – very much like me; a bit like me ; not really like me; not at all like me. Item i-iv are scored as 4,3,2, and 1, such that larger numbers are all associated with greater conscientiousness. We construct an index by computing Equation 5.

   (a) Conscientiousness index

      i. I often finish whatever I start.

      ii. I can manage to solve most difficult problems if I try hard enough.

      iii. If I am not good at a task, I can still do well on it by working hard.

      iv. I keep working to solve a problem, even if the problem is difficult.

3. **Planning**: Students can choose between four options on a likert-scale of – very much like me; a bit like me ; not really like me; not at all like me. Item i-ii are scored as 4,3,2, and 1, respectively, and items iii are reverse scored, i.e., 1-4 respectively. Such that larger numbers are all associated with greater planning behaviors. We construct an index by computing Equation 5.

   (a) Planning index

      i. I typically develop schedules and plans with clear and specific deadlines.

      ii. I usually arrange the tasks I need to do in a logical order before I start doing them.

iii. Normally, I directly start working on tasks, rather than developing a plan on how to do them.

4. **Mental experiencing**: This is a measure of student's clarity in imagining their target future career, similar to the concept of Ashraf et al. (2024). Students were asked: "Think about your targeted career, on a scale from 0% - 100%, with 100% being extremely motivated, how confident are you that you know the steps you need to take to achieve this targeted career?" They could choose between four options – less than 25%; 26% - 50%; 51% - 75%; more than 75%. We construct a continuous variable that takes the value of the midpoint of category and standardize this using the control group mean.

5. **Subjective well being**: Items are based on School Burnout Inventory (Salmela-Aro et al., 2009). Students can choose between four options on a likert-scale of – very much like me; a bit like me ; not really like me; not at all like me. Item i-iii are reverse scored, i.e., 1-4 respectively. Such that larger numbers are all associated with lower negative well-being. We construct an index by computing Equation 5.

   (a) Well-being index

      i. I feel overwhelmed by my schoolwork.

      ii. I feel a lack of motivation in my schoolwork and often think about giving up.

      iii. I sometimes sleep badly because of matters related to my schoolwork.

6. **Motivations**: Items based on Academic Self-regulation Scale (Ryan and Connell, 1989). Students were asked to choose which of the four statements describes them best regarding three questions:

   (a) classwork: students could choose between four options – I do my classwork because I want to learn new things.; I do my classwork because I will feel bad about myself

if it doesn't get done.; I do my classwork because I enjoy it.; I do my classwork because that is the rule.

(b) school: students could choose between four options – I try to do well in school because l enjoy it.; I try to do well in school so my teachers will think I am a good student.; I try to do well in school because I like doing a good job on my school work.; I try to do well in school because I will get in trouble if I don't.

(c) homework: students could choose between four options – I do my homework so that the teacher will not get upset at me.; I do my homework because I want the teacher to think I am a good student.; I do my homework because I want to learn new things.; I do my homework because it is fun.

Each question contains a choice each that maps as external regulation, introjected regulation, identified regulation and intrinsic motivation. We construct four indices for external regulation, introjected regulation, identified regulation, and intrinsic motivation by constructing a variable X_{jk}, where j is a trait, and $k \in \{1, 2, 3\}$ indexes the questions. We then construct indices for each trait j by taking the average of the z-scores of X_{jk}, across questions k.

(a) External regulation index (Extrinsic motivation)

   i. I do my classwork because that is the rule.

   ii. I try to do well in school because I will get in trouble if I don't.

   iii. I do my homework so that the teacher will not get upset at me.

(b) Introjected regulation index (Extrinsic motivation)

   i. I do my classwork because I will feel bad about myself if it doesn't get done.

   ii. I try to do well in school so my teachers will think I am a good student.

   iii. I do my homework because I want the teacher to think I am a good student.

(c) Identified regulation index (Extrinsic motivation)

i. I do my classwork because I want to learn new things.

ii. I try to do well in school because I like doing a good job on my school work.

iii. I do my homework because I want to learn new things.

(d) Intrinsic motivation index

i. I do my classwork because I enjoy it.

ii. I try to do well in school because l enjoy it.

iii. I do my homework because it is fun.

7. **Expectations about educational outcomes**

(a) Students were asked: "what marks out of 500 do you expect to get on your KCPE?" They could choose between four options – less than 200 marks; 201 - 300 marks; 301 - 400 marks; more than 400 marks. We construct a continuous variable that takes the value of the midpoint of category.

(b) Students were asked: "what type of secondary school do you expect to attend?" We construct three binary variables that take on value 1 (0 otherwise) if the student responded:

i. National level

ii. National or extra-county level

iii. National or extra-county level or county level

(c) Students were asked: "what percentage do you expect to get for Math on the upcoming end term exam?" They could choose between four options – less than 50%; 51% - 65%; 66% - 80%; more than 80%. We construct a continuous variable that takes the value of the midpoint of category.

(d) Students were asked: "what percentage do you expect to get for English on the upcoming end term exam?" They could choose between four options – less than

50%; 51% - 65%; 66% - 80%; more than 80%. We construct a continuous variable that takes the value of the midpoint of category.

8. **Homework behavior**

   (a) Students were asked: "on a usual day, how many hours do you spend on doing homework after you get home from school?" They could choose between four options – 30 minutes; 1 hour, 1 hour 30 minutes, 2 hours. We construct a continuous variable taking on the value (in minutes) of each category.