The Impact of Online Reading Material in Secondary School on University Progression*

Laura Derksen ⓡ Catherine Michaud-Leclerc

July 9, 2026

Abstract

Secondary schools in low-income countries often lack the resources to prepare students to progress to higher education. We study the long-run effects of a randomized intervention that promotes reading among Malawian students through access to online information (Wikipedia). One year of access raises later national exam scores by 0.14 standard deviations, increases the probability of qualifying for university by 7.2 percentage points, and doubles university enrollment, with effects concentrated among low achievers and girls. The effect grows over time after the intervention ends, consistent with dynamic complementarities from English proficiency.

*Derksen: Frisch Centre for Economic Research. Contact: laura.derksen@frisch.uio.no. Michaud-Leclerc: Université Laval. Contact: catherine.michaud-leclerc@ecn.ulaval.ca. We are grateful for insightful comments from Natalie Bau, Kjetil Bjorvatn, Vincent Boucher, Sue Dynarski, Alex Eble, Foster Gondwe, Alan Griffith, Jonas Hjort, Aristide Houndetoungan, Jason Kerwin, Andreas Kotsadam, Vincent Somville, and seminar and conference participants. We thank Abdallah Chilungo for project management assistance and John Adegbonmire, Stéphanie Poirier, Erling Stray and Emerick Vachon for research assistance. This study was approved by the University of Toronto Research Ethics Board, the Comités d'éthique de la recherche avec des êtres humains de l'Université Laval, and the Malawi National Committee on Research Ethics in the Social Sciences and Humanities. The study was funded by SSHRC Insight Development and Partnership Engage grants. AEA Trial Registry ID AEARCTR-0003824 (https://www.socialscienceregistry.org/trials/3824, see Analysis Plan for Long-Run Education Outcomes). The authors used Claude Code to edit and debug analysis code and for proofreading. All errors and omissions are our own. Author order randomized at https://www.aeaweb.org/journals/policies/random-author-order/search?RandomAuthorsSearch%5Bsearch%5D=-m7pwojjrMkh

1 Introduction

Higher education is a main driver of structural transformation, economic development, and poverty reduction worldwide (Porzio et al., 2022; Cox, 2025; Gethin, 2025; Lallemant and Dillon, 2026). While nearly every child now enrolls in primary school, the education gap between rich and poor countries widens in secondary school, and most dramatically in higher education. In low-income countries, the higher-education enrollment rate is less than 10 percent, compared to 79 percent in high-income countries (UNESCO, 2025).

The reasons for this gap are unclear. Relative to income, Sub-Saharan African governments spend more per higher-education student than anywhere in the world (UNESCO, 2025). The number of universities in the region has expanded from 130 in 1990 to over 1500 in 2014 (Darvas et al., 2017). The private returns are high (Patrinos and Psacharopoulos, 2020; Somani, 2021; Donath et al., 2021; Elsayed and Shirshikova, 2023; Patrinos, 2024), and scholarships and loans are often available. One important factor is learning: children across low-income countries often lack foundational skills (Angrist et al., 2021), and secondary schools may not sufficiently prepare students for higher education.

The global internet expansion creates an opportunity for schools that lack learning resources, by providing access to free information and reading material. Yet, the internet also provides distraction, misinformation, harmful content, and mass-generated AI output. Unrestricted access has typically not delivered learning gains, as young users mostly seek entertainment (Goolsbee and Guryan, 2006; Vigdor et al., 2014; Faber et al., 2015; Malamud et al., 2019; Cueto et al., 2025). Can schools support learning through restricted access to the internet's most fundamental feature: open information?

This paper shows that providing access to online information in secondary school can promote reading, improve long-run learning outcomes, and increase enrollment in higher education. We use data from an experiment that provided Wikipedia access to Malawian secondary school students. Wikipedia, the world's largest online encyclopedia, provides free reading material on nearly any topic of interest. The intervention took place during the 2017-18 school year in four academically selective public boarding schools. As is common across African countries, public boarding schools serve promising students from remote areas, and are the main path to higher education. Like other Malawian schools, the schools are under-resourced and do not allow phones or provide internet access. One-fifth of students were randomly assigned to gain access to a "digital library" where they could use restricted devices to browse Wikipedia after school and on weekends. The intervention was discontinued after one school year.

Previous work showed that the intervention effectively promotes reading and improves

English proficiency (Derksen et al., 2022). Students used Wikipedia primarily for general interest reading, with a total average reading time of 29 hours. This led to a 0.1 standard deviation increase in English scores at the end of the school year, and no significant average treatment effect on other subjects. The impact was concentrated among those with below-median scores at baseline.

In this paper, we document long-run positive impacts across subjects, with improved national final exam scores and increased university qualification, enrollment and attendance. Students who had Wikipedia access score 0.14 standard deviations higher on the national exam and are 7.2 percentage points more likely to qualify for university admission (both $p < 0.01$). These effects are concentrated among initially lower-achieving students, who score 0.26 standard deviations higher and are 13 percentage points more likely to qualify. Treatment students are more than twice as likely to enroll in university (6.4 percentage points, $p < 0.01$), and attend at least one year of university (5.3 percentage points, $p < 0.05$) within three years of secondary school. Effects on university enrollment are larger for low achievers and concentrated among girls. Indeed, families of relatively higher socioeconomic status are more likely to send lower achievers and girls to secondary school, and are more likely to be able to afford university tuition.

The effects *fade in* over time, suggesting dynamic complementarities from English proficiency as a primary mechanism (Cunha and Heckman, 2007; Bailey et al., 2020). Estimated effects grow over time, within and across cohorts, and spread from English to other subjects. The effects are concentrated among those with poorer baseline English, not poorer overall performance. In this selective school setting, even those with relatively poorer English proficiency would have performed above average on the primary exam, and therefore had the baseline English skills to read Wikipedia articles with some comprehension. Gains on the national secondary exam are concentrated in subjects taught in English, with no significant impact on Chichewa (a local language). Early access is valuable not only because gains compound over time, but also because late access appears to distract higher-achieving students in their final year, reducing their probability of university qualification. We rule out alternative primary mechanisms, including increased stock of knowledge or academic motivation. We find, using both friendship networks and nationwide panel data, that spillovers to the control group were likely limited.

The intervention delivers a particularly high return, even compared to the most cost-effective secondary school interventions, with an estimated social benefit-cost ratio of 10.9 and an infinite marginal value of public funds. The intervention cost is $26 per 0.1 standard deviation improvement in exam scores, or $554 per additional higher education enrollee. Scholarships improve outcomes for Ghanaian secondary students ($300 per 0.1 standard deviations, or \$6,200 per additional higher education enrollee; Duflo et al. 2025), and Ashraf et al. (2020, 2025) show the value of better negotiating skills for Zambian girls (\$970 per additional higher education graduate).

However, fee-based and learning-based interventions do not operate on the same margin. Like scholarships, the online reading intervention does increase higher education primarily among girls. Yet, scholarships reach students who cannot afford secondary school, many of whom would also struggle to pay for higher education, making this an expensive way to increase enrollment. Our intervention instead improves learning outcomes, and thus reaches students who could find a way to pay for higher education but fall short of qualifying. Moving this margin is cost-effective, but it does not necessarily draw the highest-*ability* students into higher education.

This study contributes to a new literature on promoting higher education in low-income countries, by demonstrating the importance of learning in secondary school. Most studies of education policy in low-income countries focus on primary school or middle school (Evans and Mendez Acosta, 2021) or on vocational training and entrepreneurship (Hicks et al., 2013; Bjorvatn et al., 2020; Brown et al., 2024; Lafortune et al., 2024). Barriers to higher education may be different: Gray-Lobe et al. (2022) find that enormous learning gains among Kenyan primary school children have only a small impact on secondary enrollment. Beyond scholarships (Duflo et al., 2025, 2026) and negotiation skills (Ashraf et al., 2020, 2025), interventions that increase secondary school attainment include career planning (Almås et al., 2025), cash transfers (Baird et al., 2011) and free primary school (Deininger, 2003; Lucas and Mbiti, 2012). In middle- and high-income countries, much of the literature has focused on financial, informational and behavioral barriers to higher education (Bettinger et al., 2012; Dynarski et al., 2021), but our results are consistent with the existing evidence on the importance of learning outcomes (Sánchez and Singh, 2018) and secondary school quality (Lavy, 2020).

This paper shows how schools can use free, open-source online resources to promote educational attainment. Lakdawala et al. (2023) show how Peruvian primary schools adapted internet use to support learning over time, and Bianchi et al. (2022) find that in China, the internet connects teachers with learners in remote areas. Wikipedia offers reading material with enough breadth and relevance to capture diverse interests; simply providing reading material will only improve outcomes if students engage with it (Glewwe et al., 2009; Borkum et al., 2012; Sabarwal et al., 2014; Falisse et al., 2019).[1] We also connect to a literature on technology that can help low-achieving students catch up by teaching at the right level

[1]Students often learn best through curiosity and self-directed inquiry (Dewey, 1938; Bruner, 1961; Freire, 1970; Rancière, 1991; Alan and Mumcu, 2024).

(Banerjee et al., 2007; Linden, 2008; Barrow et al., 2009; Muralidharan et al., 2019; Beg et al., 2022; Eames et al., 2026). Unlike context-specific programs that rely on proprietary software and teacher training, Wikipedia is broad, free, and open-source, and access can be organized through a simple after-school program.[2] Wikipedia allows learners to practice reading, in both regular and simplified English, at their own pace and based on their interests.

Finally, our results demonstrate *fade-in* effects, or dynamic complementarities, arising from language proficiency. Test-score impacts of educational interventions typically fade out within a few years (Bailey et al., 2020). In high-income countries, gains from early education have been shown to fade out and then re-emerge in adulthood (Chetty et al., 2011; Dynarski et al., 2013). More recent work has identified effective early literacy programs for low-income countries (Piper et al., 2018; Cilliers et al., 2020; Fazzio et al., 2021; Eble et al., 2021; Buhl-Wiggers et al., 2024; Andersen et al., 2026), though long-run evidence remains scarce; Stern et al. (2024) find gains that persist, and partially emerge, years after an early-grade reading program. Bailey et al. (2020) call for long-run evidence on the gains from literacy, to capture the emergence of dynamic complementarities. We find that even in adolescence, the development of language skills can affect educational trajectories.

2 Context and Experiment

2.1 Secondary and Higher Education in Malawi

In Malawi, a low-income country in sub-Saharan Africa, approximately 36 percent of children enroll in secondary school (UNESCO, 2025). Malawi has 76 public boarding secondary schools, which serve promising students from both urban and remote areas and select students based on primary school performance (de Hoop, 2010). Boarding school fees are approximately 100 USD per term, and scholarships are common. Core subjects include English, Biology, Mathematics, and Chichewa (a local language). Many elective subjects are offered, such as agriculture, geography, history, physics, and chemistry. All subjects except Chichewa are taught in English, though English is rarely spoken at home.

A standardized national exam determines whether each student receives the Malawi School Certificate of Education (MSCE) and whether they qualify for admission to public universities. Subject exams are scored between 1 (best) and 9 (worst). Those who score 8 or better pass the subject, and those who score 6 or better receive a "credit". To obtain the MSCE, students must either (i) pass 6 subjects, including English, with at least 1 credit, or (ii) pass 5 subjects, including English, with at least 3 credits. The aggregate score is

[2] Oreopoulos et al. (2026) show that feasible implementation is key to ensuring sustained use.

defined by the examinations board to be the sum of the best 6 subjects, conditional on passing English. For those with fewer than 6 scores, or failing English, the aggregate score is undefined.[3]

Only 3 percent of Malawians enroll in higher education, one of the lowest rates worldwide (UNESCO, 2025). Higher education options include six public universities and a number of private universities, as well as technical and vocational, teacher-training and nursing colleges. To qualify for admission to a public university, students must obtain at least 6 credits, and pass English.[4] In 2018, 63 percent of boarding school students qualified for university. It is rare for a day school student to attend university, and only 20 percent qualify.[5]

2.2 Intervention

Wikipedia is a free, open-source online encyclopedia, providing accurate and up-to-date information on a vast range of topics.[6] It is the largest and most-read reference work, and one of the largest collaborative projects, in history (The Economist, 2021). Content is created and edited by volunteers through open collaboration. Wikipedia's accuracy on scientific topics is comparable to an offline encyclopedia (Giles, 2005; Smith, 2020).

Wikipedia is a source of engaging and accessible reading material for secondary school students, hosting over 7 million English-language articles (Wikipedia, 2025). Articles also exist in *Simple English*, which uses simpler vocabulary. Wikipedia provides similar information to that found in secondary school textbooks, and is often more detailed. It also provides reading material on topics ranging from general relativity to pop culture, and from ancient history to modern politics.

During the 2017-18 school year, we implemented an intervention in which secondary students could access Wikipedia at school (Derksen et al., 2022). It took place in four public boarding secondary schools. Each school has approximately 500 students across Forms 1 to 4 (aged approximately 14 to 18). Although these schools select academically competitive students, fewer than 1 in 10 later enroll in higher education. Many students are of lower socioeconomic status, and nearly half do not have running water at home. Students rely on a small library and privately-owned books for reading. Phones are prohibited and internet access is not provided.

The intervention involved setting up a "digital library" in each school, equipped with Wikipedia-restricted Android devices.[7] The digital library was open, and supervised, for

[3] https://www.maneb.edu.mw/examinations/school-examinations/msce.html
[4] https://pus.nche.ac.mw/
[5] Based on national administrative data.
[6] www.wikipedia.org
[7] See Derksen et al. (2022) for an original account.

4 hours after school and 8 hours on weekends throughout the school year. Students could privately browse English Wikipedia, Simple English Wikipedia, or Wiktionary. Talking was not allowed. One in seven students could use the library at a given time, and device use was limited to 30 minutes when others were waiting. The digital libraries were closed at the end of the school year, after which students no longer had internet access.

2.3 Experimental Design

At the start of the school year, we introduced the project to students and collected baseline data. Our sample includes all 1,508 students in Forms 2–4 who consented to the study. We randomly assigned 301 students to the treatment group, and 1,207 to the control group. We also randomly selected 298 control students, who together with the treatment group form a *supplementary survey sample*. Randomization was stratified on school, form, high achiever, and past internet experience. High achievers are those with an above-median average exam score, within school and form, based on the previous year's final exams for English and Biology.[8] The randomization appears balanced across baseline variables (Table A1).

Treated students were invited to a mandatory induction session in the digital library. Each participant drew a username from one of several hats corresponding to coarsened student characteristics. It was made clear to students that browsing data collected by the research team would be anonymous. Throughout the school year, treated students could use the digital library, while others were denied entry.

We collected several types of data. Software on the Android devices captured detailed browsing data. In addition to the full-sample baseline survey, we conducted end-of-year surveys—a short version for the full sample and a longer version for the supplementary survey sample—and a 5-to-7-year follow-up survey among those in the supplementary survey sample. We collected administrative data on exam scores.

2.4 End-of-Year Results: Browsing Data and Exam Scores

In this section, we briefly describe the shorter-term findings published in Derksen et al. (2022).

Treatment students used the digital library intensively, primarily for general interest reading. The average student spent 1 hour and 20 minutes per week browsing Wikipedia, and visited 878 unique pages. Students browsed pages on a wide range of topics, including people, entertainment, music, sports, news, and sex (Figure A1). Based on a careful mapping

[8]These scores were primary short-term outcomes; we have data for both English and Biology scores for 96 percent of students at baseline.

of pages to the school syllabus, students spent 22 percent of browsing time on school-related topics.

The intervention had a positive and significant impact on end-of-year English exam scores (0.1 standard deviations), driven entirely by the effect on low achievers (0.2 standard deviations, Figure 1, Panel A). We also found a positive impact on Biology scores for low achievers only (0.14 standard deviations), and did not find significant impacts on other subjects.

3 Data and Empirical Strategy

3.1 Administrative Data

To capture national final exam performance and university qualification at the student level, we rely on administrative data from the schools from years 2018, 2019 and 2020, corresponding to the three cohorts in the study (Figure A2). We also use administrative data on school-level final exams from the 2016-17 school year for baseline measures, and we use nationwide panel data from 2015–2024 to estimate spillovers (see details in Appendix B).

We construct primary outcomes including exam scores, whether the student passed, and whether they qualified for university admission. First, we reverse all scores so higher values indicate better performance. Then, we normalize each core subject grade, subtracting the control-group mean and dividing by the control-group standard deviation.[9] We also construct a normalized MSCE aggregate score as defined by the Malawi National Examinations Board. We construct an indicator for passing secondary school, and an indicator for whether the student qualified for admission to public universities (see Section 2.1). Not all students sit all subject exams, and some scores are therefore missing, with no significant difference by treatment status (see Panel A of Table A2). Indicators for pass and qualification are set to 0 for all students who did not pass or qualify with their cohort.

3.2 Survey Data

This paper also uses data from baseline, end-of-year, and long-term surveys. The baseline survey was administered to 1,508 students, who form the full sample. The long-term survey took place between 2023 and 2025, when all cohorts should have graduated at least three years earlier. We attempted to survey all 301 treated students as well as the 298 control students in the supplementary survey sample. We contacted respondents using the details they provided in the end-of-year survey and through social networks, and included strict identity

[9]In our primary analysis we normalize with respect to the full control group to facilitate comparison across cohorts.

checks. Most interviews took place over the phone, though 11 surveys were conducted in person.[10] We achieved an 85.3 percent response rate and do not find differential attrition by treatment status (see Panel B of Table A2).

We use this data to construct additional higher education and employment outcomes within three years of secondary school. We use retrospective questions and truncate at three years to construct comparable outcomes for all three cohorts. We classify these outcomes as secondary, as they rely on recall data, and the sample is smaller than that observed in administrative data.

3.3 Empirical Strategy

Our primary analysis estimates the following specification:

$$y_i = \beta^L \text{Treatment}_i \times \text{LowAchiever}_i + \beta^H \text{Treatment}_i \times \text{HighAchiever}_i + \boldsymbol{\chi}\mathbf{X}_i + \sigma_s + \varepsilon_i \quad (1)$$

Here, y_i is the outcome for individual i, Treatment_i is an indicator for the treatment group, LowAchiever_i and HighAchiever_i are indicators for having a below- or above-median final exam score in the year prior to the intervention, as defined in Section 2.3. The two interaction terms partition the treatment group; we therefore omit a treatment indicator, and β^L and β^H capture the treatment effect for each achievement group. Covariates $\mathbf{X}_i$ include English exam score, normalized within school and form, in the year prior to the intervention, and an indicator for missing score. σ_s are indicators for the stratification bins, which are collinear with HighAchiever_i. ε_i is an error term. We use ordinary least squares, with robust standard errors, to estimate this equation on two separate subsamples: those for whom the intervention took place in the same year as graduation (Form 4) and earlier-intervention cohorts (pooling Forms 2 and 3). We report both conventional and randomization-inference p-values.

We registered primary outcomes and analyses for the long-run study after analyzing short-term data but before analyzing long-term data. The original pre-registered outcomes for the short-term study were English and Biology exam scores, which we expanded to include other core subjects, and indicators for passing and qualifying for university. Heterogeneity by baseline achievement was pre-specified for both short and long-run analysis, to assess the intervention's potential to close achievement gaps. Because short-run learning gains were concentrated among low achievers (Derksen et al., 2022), we hypothesized that long-run effects would follow the same pattern. In addition to the registered analysis, we estimate a specification pooling students across cohorts, average treatment effects (without

[10] In-person surveys were balanced across treatment (6 treated and 5 control).

interacting Treatment_i with HighAchiever_i and LowAchiever_i), and heterogeneous effects by socioeconomic status and gender. We did not pre-register survey-based secondary outcomes as we were initially uncertain whether we could contact participants after 5 to 7 years. In the end, survey attrition was only 15 percent. We adjust for multiple pre-registered hypotheses using FDR sharpened q-values (Anderson, 2008).

4 Results

4.1 National Secondary School Exam and University Qualification

We find that one year of access to Wikipedia increases national secondary school exam scores and likelihood of qualification for university. Treatment students receive a higher aggregate score (0.14 standard deviations, $p < 0.01$), and are 7.2 percentage points more likely to qualify for university ($p < 0.01$, Panel A of Table 1).[11] Effects are concentrated among low achievers, whose aggregate scores are 0.26 standard deviations higher ($p < 0.01$), and who are 13 percentage points ($p < 0.01$) more likely to qualify for university than their control peers. Treated low achievers have higher scores on core English-language subject exams (English, Mathematics and Biology), and on aggregate (which includes English-language electives), but not on the Chichewa exam. We find no significant impact on the probability of passing.

Pre-registered hypotheses. We hypothesized positive impacts across outcomes in Table 1 for low achievers, in both earlier-intervention cohorts and the same-year cohort, and no effect for high achievers. The impact on university qualification is significant ($q < 0.05$) for both cohorts when adjusted for multiple tests. The impact on aggregate score is significant for earlier-intervention cohorts ($q < 0.01$), and for the same-year cohort with $q < 0.1$.

The effect on final exam score appears to *fade in* rather than fade out. Although earlier-treated cohorts lost access to Wikipedia one or two years before graduation, they score 0.19 standard deviations higher on the final exam ($p < 0.01$). This effect is much larger than the estimate for those who received the intervention in their final year (0.036, with the difference significant at $p < 0.1$). For low achievers, the effect is positive and significant for both subsamples, but the estimate is approximately twice as large for earlier-treated cohorts (Panels B and C of Table 1). For high achievers, results point in opposite directions. In the final year the intervention may serve as a distraction: most estimates are negative, including an 11 percentage point decrease in university qualification ($p < 0.1$, Panel B of Table 1).[12] High achievers in the earlier-treated cohorts see a positive but insignificant

[11] See Figure 2 for score distributions. Within subjects, treatment students are more likely to exceed the credit threshold required for qualification (Table A3).

[12] While estimated on a small subsample, this effect does not appear to be due to baseline imbalance.

increase in the aggregate exam score, and a significant 6.3 percentage point increase in university qualification ($p < 0.05$, Panel C).

We also see the effect grow *over time within* the early-intervention cohorts (Figure 1, Panel A). For low achievers in these cohorts, the impact on English is large in the first year (0.2 standard deviations), and grows slightly over time, while the aggregate effect grows significantly ($p < 0.05$) over time, from 0.02 to 0.19 standard deviations.[13]

4.2 University Enrollment and Attendance

We find a positive impact on university enrollment and attendance within three years of secondary school (Panel A of Table 2). Treatment respondents are 6.5 percentage points more likely to enroll in, and 5.6 percentage points more likely to attend at least one year of, higher education (including universities, colleges, and other postsecondary institutions, $p < 0.05$ for both estimates). They are 6.4 percentage points more likely to enroll in a university program specifically ($p < 0.01$). Estimated effects are approximately three times larger for low achievers compared to high achievers (for whom estimated effects are also positive), but the difference is not statistically significant. Effects are fairly similar across cohorts (Table A4), and are driven by enrollment in accredited and private universities (Table A5). In the end-of-year survey, both treatment and control students had high career aspirations (Figure A3), though most do not enroll in their first-choice programs.

We find no significant effect on employment, though point estimates for low achievers are negative, suggesting a shift towards education (Column 5 of Table 2). Paid employment is rare both in our sample (6 percent) and among 18–21-year-old secondary-educated Malawians (4 percent, Malawi IHS5 2019), as many engage in casual or unpaid labor.

4.3 Heterogeneity by Socioeconomic Status and Gender

The impact on university enrollment is entirely concentrated among girls. Treated girls are 13 percentage points ($p < 0.01$) more likely to enroll in university than control girls, only 3 percent of whom enroll (Panel B of Table 2). This represents a difference in enrollment behavior rather than a differential impact on learning; the impact on university qualification is similar for boys and girls (Panel A of Table 3). The intervention goes beyond closing the gender gap: treated girls are more than twice as likely to go to university as treated boys.

The large impacts on university enrollment for girls and low achievers may be partly explained by ability to pay: girls are 22 percentage points, and low achievers 7 percentage

[13] Here, we use the specification from Derksen et al. (2022), normalizing within forms and schools, to compare school exam scores to national final exams.

points, more likely to be of higher socioeconomic status (SES, having running water and electricity at home). Indeed, poorer families are more likely to send boys and higher-achieving children to boarding secondary schools. For low achievers, the increase in university enrollment approximately matches the increase in qualification (Figure 2, Panel B). High achievers are more likely to qualify, but low achievers are approximately equally likely to enroll, even in the control group. While Wikipedia had a similar effect on qualification for students of both higher and lower SES (Panel B of Table 3), the estimated impact on enrollment is much larger for higher-SES students (though the difference is not significant, Panel C of Table 2).[14]

4.4 Robustness

Lee (2009) bounds, reported in Tables A10, A11 and A12 for national exam and survey outcomes, account for potential bias from attrition. Passing and university qualification are observed for all students and unaffected by attrition. Attrition is low, and not differential by treatment status, for both aggregate MSCE scores and survey outcomes (Panels A and B of Table A2). The bounds are therefore quite narrow and close to main effect estimates, though imprecisely estimated for survey outcomes. National exam results are also similar using raw exam scores (Table A13),[15] and in the surveyed subsample (Table A14).

4.5 Mechanisms

Why do academic effects *fade in* over time? We shed light on an important mechanism: dynamic complementarities from English proficiency. We also rule out stock-of-knowledge, motivation, and network channels as primary mechanisms.

Dynamic complementarities from English language skills. Bailey et al. (2020) highlight English language proficiency as a *fundamental skill* — it is malleable, built upon, and affects later learning and labor market outcomes. It generates dynamic complementarities (Cunha and Heckman, 2007; Bailey et al., 2020): early gains in English raise the productivity of subsequent schooling, helping students who speak English as a second language understand teachers, books and exams. Indeed, even Mathematics exams involve complicated word problems (see Appendix D).

Several empirical findings point to this channel. During the intervention year, treatment students spent 29 additional hours, on average, reading in English, while their control-group peers engaged primarily in recreation (Derksen et al., 2022). This durably increases English scores (Figure 1, Panel A). Over time, the impacts spread to other subjects, impacting the

[14]See Tables A6 to A9 for results by cohort.

[15]Normalization, while not pre-specified, does not affect inference.

aggregate final exam score, but not Chichewa (the only non-English subject). The effect on academic outcomes grows through a spread to other subjects, and not through a significant increase in English scores over the longer term. All estimated effects are concentrated among low achievers, and specifically among students with poor baseline English scores (Figure 1, Panels B and C).

Stock-of-knowledge from a better study tool. While students did visit syllabus-related Wikipedia pages, most of this time was spent on Biology (Derksen et al., 2022). This can partially explain the impact on Biology, but not on other subjects. For example, we find an impact on Mathematics for low achievers, though the average low achiever spent only 14 minutes reading Mathematics-related pages. Moreover, knowledge tied to the syllabus during the intervention year would be expected to depreciate as the syllabus moves on. A stock-of-knowledge channel cannot easily explain effects that grow after access ends, nor their concentration among students with poor baseline English.

Academic motivation. The intervention does not appear to work by raising academic motivation (or by demotivating control students). Treatment and control students report similar study time, career aspirations, ambition, confidence, and happiness (Derksen et al., 2022). Moreover, if the effect were driven by increasing academic motivation, we would expect a positive long-run impact on those with decent English but low baseline scores in other subjects, and this is not the case (Figure 1, Panel C). Instead, the intervention appears to boost those who have the motivation and financial ability to attend university, but fall behind due to poor English.

Spillovers and networks. Control students were prevented from accessing Wikipedia directly. They were denied entry to the digital library, with identities verified using photographs, and research staff conducted spot checks.

We must nevertheless consider indirect spillovers. Students may have benefited from information sharing between peers, or from having peers with better English. Yet, a specification using baseline networks found small and insignificant spillovers during the intervention year (Derksen et al., 2022). This is perhaps unsurprising given the main mechanism. English proficiency likely improves with direct exposure to reading material rather than through information sharing.

We expand this analysis to estimate long-run spillovers using baseline network data, an endogenous network model, and nationwide administrative panel data on exam performance. The analyses are described in detail in Appendix B. We find that having more treated friends at baseline has a negligible and insignificant impact on both aggregate exam scores and qualification for university, for both control and treatment students (Table A15).[16] The

[16]While the network is not static (Derksen and Souza, 2025), the number of treated friends at baseline

estimates are also stable under a model that allows for endogenous network change and endline network spillovers (Table A16, Comola and Prina 2021). Finally, we use two-way fixed effects to compare outcomes among exam-takers in the four study schools to those from other boarding schools across Malawi and in other years (Table A17). While imprecise, the estimated impact of the intervention on control students is insignificant and close to zero: we find a 0.02 standard deviation decrease in aggregate exam score, and a 1 percentage point increase in qualification for university. The estimated impact on treated students is approximately 6 times larger for both outcomes.

Finally, the intervention itself affected information-sharing networks for both low and high achievers (Derksen and Souza, 2025), which could partially explain the long-run increase in academic scores. However, the new links were exclusively information links; personal friendship networks were not affected. Network effects might therefore weaken after students lose access to online information, in which case they could not explain the increase in academic attainment over time. Denser networks may amplify gains in English proficiency, though students primarily socialize in Chichewa. Networks do not appear to drive subject-specific improvements: end-of-year impacts were concentrated in English (Derksen et al., 2022), and, again, we do not see any positive longer-run treatment effect on those with strong English but poor aggregate scores at baseline (Figure 1, Panel C). The Comola and Prina (2021) model provides a more direct test of these types of network effects, allowing learning to spread both through baseline friendships and through new links. Both channels are estimated to be small and insignificant for all subjects, and the direct effect of treatment is essentially unchanged when both are included (Table A16).

4.6 Cost-Effectiveness

Our findings imply a cost of $26 per 0.1 standard deviation increase in exam score, or $554 per additional higher education enrollee within three years of secondary school (all detailed calculations in Appendix C). This compares favorably to secondary school interventions with the highest documented returns (Table A18).

We follow the framework of Hendren and Sprung-Keyser (2020) to calculate the benefit-cost ratio (BCR) and the marginal value of public funds (MVPF) associated with the intervention and its impact on higher education. While we cannot observe labor market effects at this point, we predict a long-run earnings effect based on a causal estimate from Ghana (Duflo et al., 2025). We apply this effect to the earnings profile of secondary-educated Malawians in Malawi IHS5 (2019), including formal, informal, and imputed farm-labor income, over a

strongly predicts the number at endline (0.4, $p < 0.001$).

working life from age 32 to 60, discounted at 5 percent. We calculate a complete set of private and public costs including higher education tuition, public expenditure, and forgone earnings.

We find a BCR of 10.9 and an infinite MVPF: tax revenue more than offsets public costs. Both are robust to alternative assumptions about earnings returns and wage growth (Table A19). While high, these estimates are in line with other cost-effective human capital investments (Hendren and Sprung-Keyser, 2020). Duflo et al. (2026) report a benefit-cost ratio of 3.4 and an MVPF of 4.1 for secondary scholarships in Ghana, and Ashraf et al. (2025) a benefit-cost ratio of 7.8 and an MVPF of 11.1 for negotiation training in Zambia. Our estimates may be conservative, as the higher-education earnings premium is substantially larger in Malawi than in Ghana, and we do not include intergenerational benefits as in Duflo et al. (2026).

5 Conclusion

We find that online reading in adolescence can improve learning and alter educational trajectories for students in a low-income country. We study a randomized intervention that gave Malawian secondary school students access to Wikipedia during one school year. Early gains in English persist and spread to other subjects over time: treated students perform better on the national final exam, and are more likely to qualify for and enroll in university. Gains are concentrated among students with lower English proficiency, and are restricted to subjects taught in English, consistent with dynamic complementarities as the primary mechanism.

The results suggest that restricted internet access can be an effective policy tool. Wikipedia is free and available in over 300 languages, and offers engaging reading material at an appropriate level for secondary students with some baseline proficiency. The intervention is cost-effective, and requires no teacher training or curriculum adaptation. The fact that effects are largest for lower-achieving students (within selective public schools) and for girls points to a potential to narrow gaps in educational attainment. On the other hand, unlike scholarships, this learning-based secondary intervention promotes higher education primarily among those who are more likely to afford fees but fall short of qualifying.

One open question is how the effects scale when access is provided at the school level. Malawian teachers may adapt their instruction to support independent inquiry (Gondwe, 2024), or reduce effort. If returns to higher education are indeed large, the university enrollment effects we document will have meaningful economic consequences, but causal evidence is needed.

As the internet expands and evolves, there is a need for evidence-based policy on internet in schools. Which parts of the internet promote learning, and under what conditions? These questions are especially relevant in post-primary education, which is under-studied, and where students require deep subject-specific knowledge. Wikipedia captures one of the internet's fundamentally educational features: open information. Online communities may support peer learning, but also expose students to distraction and misinformation. Artificial intelligence offers personalized explanations, but raises concerns about dependence and accuracy. Research is needed on how the internet's best features can be protected and harnessed for learning.

References

Alan, S. and Mumcu, I. (2024). Nurturing childhood curiosity to enhance learning: Evidence from a randomized pedagogical intervention. *American Economic Review*, 114(4):1173–1210.

Almås, I., Caeyers, B., Dautheville, A., Kazi, V., Krutikova, S., and Somville, V. (2025). Reality bites: Experimental evidence on the transition from school in a low-income setting. Discussion Paper SAM 19/25, NHH Norwegian School of Economics, Department of Economics.

Andersen, E. T. J., Graffy, S., Kerwin, J. T., and Lambon-Quayefio, M. (2026). How to build a reader: Evidence from a scalable literacy intervention in Ghana. Working Paper.

Anderson, M. L. (2008). Multiple inference and gender differences in the effects of early intervention: A reevaluation of the Abecedarian, Perry Preschool, and early training projects. *Journal of the American Statistical Association*, 103(484):1481–1495.

Angrist, N., Djankov, S., Goldberg, P. K., and Patrinos, H. A. (2021). Measuring human capital using global learning data. *Nature*, 592(7854):403–408.

Ashraf, N., Bau, N., Low, C., and McGinn, K. (2020). Negotiating a better future: How interpersonal skills facilitate intergenerational investment. *The Quarterly Journal of Economics*, 135(2):1095–1151.

Ashraf, N., Bau, N., Low, C., and Shan, X. (2025). Ten years of relational power: The long-run effects of teaching negotiation skills to adolescent girls. Working Paper 34339, National Bureau of Economic Research.

Athey, S., Chetty, R., Imbens, G. W., and Kang, H. (2019). The surrogate index: Combining short-term proxies to estimate long-term treatment effects more rapidly and precisely. Working Paper 26463, National Bureau of Economic Research.

Bailey, D. H., Duncan, G. J., Cunha, F., Foorman, B. R., and Yeager, D. S. (2020). Persistence and fade-out of educational-intervention effects: Mechanisms and potential solutions. *Psychological Science in the Public Interest*, 21(2):55–97.

Baird, S., McIntosh, C., and Özler, B. (2011). Cash or condition? Evidence from a cash transfer experiment. *The Quarterly Journal of Economics*, 126(4):1709–1753.

Banerjee, A. V., Cole, S., Duflo, E., and Linden, L. L. (2007). Remedying education: Evidence from two randomized experiments in India. *The Quarterly Journal of Economics*, 122(3):1235–1264.

Barrow, L., Markman, L., and Rouse, C. E. (2009). Technology's edge: The educational benefits of computer-aided instruction. *American Economic Journal: Economic Policy*, 1(1):52–74.

Beg, S., Halim, W., Lucas, A. M., and Saif, U. (2022). Engaging teachers with technology increased achievement, bypassing teachers did not. *American Economic Journal: Economic Policy*, 14(2):61–90.

Bettinger, E. P., Long, B. T., Oreopoulos, P., and Sanbonmatsu, L. (2012). The role of application assistance and information in college decisions: Results from the H&R Block FAFSA experiment. *The Quarterly Journal of Economics*, 127(3):1205–1242.

Bianchi, N., Lu, Y., and Song, H. (2022). The effect of computer-assisted learning on students' long-term development. *Journal of Development Economics*, 158:102919.

Bjorvatn, K., Cappelen, A. W., Sekei, L. H., Sørensen, E. Ø., and Tungodden, B. (2020). Teaching through television: Experimental evidence on entrepreneurship education in Tanzania. *Management Science*, 66(6).

Borkum, E., He, F., and Linden, L. L. (2012). The effects of school libraries on language skills:

Evidence from a randomized controlled trial in India. Working Paper 18183, National Bureau of Economic Research.

Brown, G., Hardy, M., Mbiti, I., McCasland, J., and Salcher, I. (2024). Can financial incentives to firms improve apprenticeship training? Experimental evidence from Ghana. *American Economic Review: Insights*, 6(1):120–136.

Bruner, J. S. (1961). The act of discovery. *Harvard Educational Review*, 31:21–32.

Buhl-Wiggers, J., Kerwin, J. T., Muñoz, J. S., Smith, J. A., and Thornton, R. L. (2024). Some children left behind: Variation in the effects of an educational intervention. *Journal of Econometrics*, 243(1):105256.

Chetty, R., Friedman, J. N., Hilger, N., Saez, E., Schanzenbach, D. W., and Yagan, D. (2011). How does your kindergarten classroom affect your earnings? Evidence from Project STAR. *The Quarterly Journal of Economics*, 126(4):1593–1660.

Cilliers, J., Fleisch, B., Prinsloo, C., and Taylor, S. (2020). How to improve teaching practice? An experimental comparison of centralized training and in-classroom coaching. *Journal of Human Resources*, 55(3):926–962.

Comola, M. and Prina, S. (2021). Treatment effects accounting for network changes. *Review of Economics and Statistics*, 103(3):597–604.

Cox, A. (2025). Fostering development through higher education: College attainment, firms and economic growth. Working Paper, Universidad Carlos III de Madrid.

Cueto, S., Beuermann, D. W., Cristia, J., Malamud, O., and Pardo, F. (2025). Laptops in the long run: Evidence from the one laptop per child program in rural Peru. *Journal of Public Economics*, 252:105538.

Cunha, F. and Heckman, J. (2007). The technology of skill formation. *American Economic Review*, 97(2):31–47.

Darvas, P., Gao, S., Shen, Y., and Bawany, B. (2017). *Sharing Higher Education's Promise beyond the Few in Sub-Saharan Africa*. Directions in Development—Human Development. World Bank, Washington, DC.

de Hoop, J. (2010). Selective secondary education and school participation in Sub-Saharan Africa: Evidence from Malawi. Discussion Paper 10-041/2, Tinbergen Institute.

Deininger, K. (2003). Does cost of schooling affect enrollment by the poor? Universal primary education in Uganda. *Economics of Education Review*, 22(3):291–305.

Derksen, L., Michaud-Leclerc, C., and Souza, P. C. (2022). Restricted access: How the internet can be used to promote reading and learning. *Journal of Development Economics*, 155:102810.

Derksen, L. and Souza, P. C. (2025). Who knows? Information access and endogenous network formation. Working Paper.

Dewey, J. (1938). *Logic: The Theory of Inquiry*. Holt.

Donath, L., Morrissey, O., and Owens, T. (2021). Does the pay period matter in estimating returns to schooling? Evidence from East Africa. Technical report, CREDIT Research Paper.

Duflo, E., Dupas, P., and Kremer, M. (2025). The impact of secondary school subsidies on career trajectories in a dual labor market: Experimental evidence from Ghana. Working paper.

Duflo, E., Dupas, P., Spelke, E., and Walsh, M. P. (2026). Intergenerational impacts of secondary education: Experimental evidence from Ghana. Forthcoming, American Economic Review.

Dynarski, S., Hyman, J., and Schanzenbach, D. W. (2013). Experimental evidence on the effect of childhood investments on postsecondary attainment and degree completion. *Journal of Policy Analysis and Management*, 32(4):692–717.

Dynarski, S., Libassi, C. J., Michelmore, K., and Owen, S. (2021). Closing the gap: The effect of reducing complexity and uncertainty in college pricing on the choices of low-income students. *American Economic Review*, 111(6):1721–1756.

Eames, T., Brunskill, E., Yamkovenko, B., Weatherholtz, K., and Oreopoulos, P. (2026). Computer-assisted learning in the real world: How Khan Academy influences student math learning. *Proceedings of the National Academy of Sciences*, 123(1):e2507708123.

Eble, A., Frost, C., Camara, A., Bouy, B., Bah, M., Sivaraman, M., Hsieh, P.-T. J., Jayanty, C., Brady, T., Gawron, P., Vansteelandt, S., Boone, P., and Elbourne, D. (2021). How much can we remedy very low learning levels in rural parts of low-income countries? Impact and generalizability of a multi-pronged para-teacher intervention from a cluster-randomized trial in the Gambia. *Journal of Development Economics*, 148:102539.

Elsayed, A. and Shirshikova, A. (2023). The women-empowering effect of higher education. *Journal of Development Economics*, 163:103101.

Evans, D. K. and Mendez Acosta, A. (2021). Education in Africa: What are we learning? *Journal of African Economies*, 30(1):13–54.

Faber, B., Sanchis-Guarner, R., and Weinhardt, F. (2015). ICT and education: Evidence from student home addresses. Working Paper 21306, National Bureau of Economic Research.

Falisse, J.-B., Huysentruyt, M., and Olofsgård, A. (2019). Incentivizing textbooks for self-study: Experimental evidence on student learning from the Democratic Republic of Congo. Working Paper.

Fazzio, I., Eble, A., Lumsdaine, R. L., Boone, P., Bouy, B., Hsieh, P.-T. J., Jayanty, C., Johnson, S., and Silva, A. F. (2021). Large learning gains in pockets of extreme poverty: Experimental evidence from Guinea Bissau. *Journal of Public Economics*, 199:104385.

Freire, P. (1970). *Pedagogy of the Oppressed*. Seabury Press.

Gethin, A. (2025). Distributional growth accounting: Education and the reduction of global poverty, 1980–2019. *The Quarterly Journal of Economics*, 140(4):2571–2618.

Ghana GLSS 7 (2017). Ghana Living Standards Survey Round 7 (GLSS 7) 2016–2017. Survey report and microdata, Ghana Statistical Service, Accra, Ghana. Ghana Statistical Service. Fieldwork October 2016–October 2017.

Giles, J. (2005). Special report internet encyclopaedias go head to head. *Nature*, 438(15):900–901.

Glewwe, P., Kremer, M., and Moulin, S. (2009). Many children left behind? Textbooks and test scores in Kenya. *American Economic Journal: Applied Economics*, 1(1):112–35.

Gondwe, F. (2024). Research-based teacher preparation for technology-enhanced education in Malawi: A case study. In *Global Perspectives on Teaching with Technology*, pages 127–141. Routledge.

Goolsbee, A. and Guryan, J. (2006). The impact of internet subsidies in public schools. *The Review of Economics and Statistics*, 88(2):336–347.

Gray-Lobe, G., Keats, A., Kremer, M., Mbiti, I., and Ozier, O. (2022). Can education be standardized? Evidence from Kenya. BFI working paper no. 2022-168, Becker Friedman Institute for Economics, University of Chicago.

Hendren, N. and Sprung-Keyser, B. (2020). A unified welfare analysis of government policies. *The Quarterly Journal of Economics*, 135(3):1209–1318.

Hicks, J. H., Kremer, M., Mbiti, I., and Miguel, E. (2013). Vocational education in Kenya: Evidence from a randomized evaluation among youth. *Nashville, TN: Vanderbilt University.*

Lafortune, J., Pugatch, T., Tessada, J., and Ubfal, D. (2024). Can gamified online training make high school students more entrepreneurial? Experimental evidence from Rwanda. *Economics of Education Review*, 101:102559.

Lakdawala, L. K., Nakasone, E., and Kho, K. (2023). Dynamic impacts of school-based internet access on student learning: Evidence from Peruvian public primary schools. *American Economic Journal: Economic Policy*, 15(4):222–254.

Lallemant, T. and Dillon, B. (2026). Beyond the campus: Indirect effects of university openings in Uganda. *AEA Papers and Proceedings*, 116:201–205.

Lavy, V. (2020). Teachers' pay for performance in the long-run: The dynamic pattern of treatment effects on students' educational and labour market outcomes in adulthood. *The Review of Economic Studies*, 87(5):2322–2355.

Lee, D. S. (2009). Training, wages, and sample selection: Estimating sharp bounds on treatment effects. *The Review of Economic Studies*, 76(3):1071–1102.

Linden, L. L. (2008). Complement or substitute? The effect of technology on student achievement in India. Abdul Latif Jameel Poverty Action Lab Working Paper.

Lucas, A. M. and Mbiti, I. M. (2012). Access, sorting, and achievement: The short-run effects of free primary education in Kenya. *American Economic Journal: Applied Economics*, 4(4):226–253.

Malamud, O., Cueto, S., Cristia, J., and Beuermann, D. W. (2019). Do children benefit from Internet access? Experimental evidence from Peru. *Journal of Development Economics*, 138:41–56.

Malawi IHS5 (2019). Fifth Integrated Household Survey (IHS5) 2019–2020. Survey report and microdata, National Statistical Office (Malawi), Zomba, Malawi. National Statistical Office (Malawi). World Bank LSMS-ISA. Fieldwork April 2019–April 2020.

Malawi Revenue Authority (2018). New income tax rates. Effective 1 July 2018.

Muralidharan, K., Singh, A., and Ganimian, A. J. (2019). Disrupting education? Experimental evidence on technology-aided instruction in India. *American Economic Review*, 109(4):1426–1460.

Oreopoulos, P., Keyes-Krysakowski, O., and Agarwal, D. (2026). How in-school supervised Ed-Tech support produces massive learning gains: A Khan Academy field experiment in India. Technical report, National Bureau of Economic Research.

Patrinos, H. A. (2024). Estimating the return to schooling using the Mincer equation. *IZA World of Labor*.

Patrinos, H. A. and Psacharopoulos, G. (2020). Returns to education in developing countries. In Bradley, S. and Green, C., editors, *The Economics of Education*, chapter 4, pages 53–64. Academic Press, second edition edition.

Piper, B., Destefano, J., Kinyanjui, E. M., and Ong'ele, S. (2018). Scaling up successfully: Lessons from Kenya's Tusome national literacy program. *Journal of Educational Change*, 19(3):293–321.

Porzio, T., Rossi, F., and Santangelo, G. (2022). The human side of structural transformation. *American Economic Review*, 112(8):2774–2814.

Rancière, J. (1991). *The Ignorant Schoolmaster*. Stanford University Press.

Sabarwal, S., Evans, D. K., and Marshak, A. (2014). The permanent input hypothesis: The case of textbooks and (no) student learning in Sierra Leone. Policy Research Working Paper 7021, World Bank.

Sánchez, A. and Singh, A. (2018). Accessing higher education in developing countries: Panel data analysis from India, Peru, and Vietnam. *World Development*, 109:261–278.

Smith, D. A. (2020). Situating wikipedia as a health information resource in various contexts: A scoping review. *PloS one*, 15(2):e0228786.

Somani, R. (2021). The returns to higher education and public employment. *World Development*, 144:105471.

Stern, J. M. B., Jukes, M. C. H., Cilliers, J., Fleisch, B., Taylor, S., and Mohohlwane, N. (2024). Persistence and emergence of literacy skills: Long-term impacts of an effective early grade reading intervention in South Africa. *Journal of Research on Educational Effectiveness*.

Tauchmann, H. (2014). Lee (2009) treatment-effect bounds for nonrandom sample selection. *The Stata Journal*, 14(4):884–894.

The Economist (2021). Wikipedia is 20, and its reputation has never been higher. *The Economist*. Archived from the original on January 7, 2021. Retrieved February 25, 2021.

UNESCO (2025). UIS statistics database. https://uis.unesco.org/. Dataset.

UNICEF (2025). Education budget brief 2025/26: Malawi. Technical report, UNICEF Malawi.

Vigdor, J. L., Ladd, H. F., and Martinez, E. (2014). Scaling the digital divide: Home computer technology and student achievement. *Economic Inquiry*, 52(3):1103–1119.

Wikipedia (2025). Wikipedia:Statistics. https://en.wikipedia.org/wiki/Wikipedia:Statistics. Accessed October 21, 2025.

World Bank (2024). World development indicators – Malawi. https://data.worldbank.org. Indicators: NY.GDP.MKTP.CD (GDP), NY.GDP.MKTP.KD.ZG (GDP growth), NY.GNP.PCAP.CD (GNI per capita, Atlas method), SE.XPD.TOTL.GD.ZS (Government expenditure on education).

Young, A. (2019). Channeling Fisher: Randomization tests and the statistical insignificance of seemingly significant experimental results. *The Quarterly Journal of Economics*, 134(2):557–598.

Tables and Figures

Table 1: National Exam Scores

	(1) English	(2) Biology	(3) Maths	(4) Chichewa	(5) Aggregate	(6) Pass	(7) Qualified for University
Panel A: All Students							
Treatment	0.085**	0.125***	0.090	0.020	0.136***	0.019	0.072***
	(0.042)	(0.045)	(0.055)	(0.044)	(0.043)	(0.017)	(0.025)
	p=0.038	p=0.003	p=0.088	p=0.650	p=0.002	p=0.300	p=0.003
Treatment × low achiever	0.213***	0.231***	0.240***	0.078	0.260***	0.044	0.130***
	(0.064)	(0.069)	(0.077)	(0.065)	(0.065)	(0.027)	(0.038)
	p=0.001	p=0.001	p=0.001	p=0.246	p=0.000	p=0.140	p=0.000
Treatment × high achiever	-0.053	0.012	-0.070	-0.042	0.006	-0.009	0.005
	(0.050)	(0.054)	(0.078)	(0.059)	(0.053)	(0.021)	(0.029)
	p=0.269	p=0.822	p=0.349	p=0.447	p=0.903	p=0.657	p=0.869
Mean control	0.000	0.000	0.000	0.000	0.000	0.892	0.701
Low achievers	-0.349	-0.514	-0.567	-0.344	-0.518	0.840	0.526
High achievers	0.360	0.530	0.581	0.349	0.530	0.951	0.896
P-value (low = high)	0.001	0.013	0.005	0.170	0.002	0.122	0.009
Observations	1381	1384	1377	1354	1378	1508	1508
Panel B: Same-year intervention (Form 4)							
Treatment	0.049	0.005	0.139*	-0.028	0.036	0.005	0.029
	(0.067)	(0.069)	(0.084)	(0.078)	(0.061)	(0.027)	(0.045)
	p=0.454	p=0.944	p=0.085	p=0.722	p=0.542	p=0.863	p=0.484
Treatment × low achiever	0.106	0.135	0.321***	0.020	0.169**	0.061*	0.153**
	(0.102)	(0.098)	(0.117)	(0.109)	(0.085)	(0.034)	(0.066)
	p=0.302	p=0.167	p=0.007	p=0.861	p=0.070	p=0.155	p=0.024
Sharpened q-value	[0.142]	[0.114]	[0.018]	[0.323]	[0.059]	[0.073]	[0.035]
Treatment × high achiever	-0.019	-0.150	-0.073	-0.084	-0.118	-0.059	-0.111*
	(0.084)	(0.091)	(0.117)	(0.113)	(0.084)	(0.041)	(0.056)
	p=0.802	p=0.102	p=0.494	p=0.427	p=0.094	p=0.049	p=0.014
Mean control	0.398	-0.287	-0.012	0.152	0.002	0.935	0.749
Low achievers	-0.001	-0.811	-0.555	-0.212	-0.500	0.901	0.575
High achievers	0.839	0.292	0.580	0.544	0.555	0.973	0.946
P-value (low = high)	0.342	0.034	0.017	0.505	0.017	0.025	0.002
Observations	474	478	474	451	477	496	496
Panel C: Earlier intervention (Forms 2 and 3)							
Treatment	0.101*	0.187***	0.065	0.042	0.189***	0.026	0.092***
	(0.053)	(0.057)	(0.072)	(0.053)	(0.056)	(0.022)	(0.029)
	p=0.052	p=0.001	p=0.346	p=0.427	p=0.001	p=0.284	p=0.002
Treatment × low achiever	0.264***	0.280***	0.194*	0.106	0.310***	0.035	0.118**
	(0.082)	(0.093)	(0.101)	(0.081)	(0.089)	(0.036)	(0.047)
	p=0.001	p=0.003	p=0.041	p=0.207	p=0.000	p=0.373	p=0.013
Sharpened q-value	[0.009]	[0.011]	[0.059]	[0.119]	[0.008]	[0.149]	[0.027]
Treatment × high achiever	-0.067	0.091	-0.068	-0.022	0.069	0.015	0.063**
	(0.062)	(0.065)	(0.101)	(0.069)	(0.066)	(0.024)	(0.031)
	p=0.290	p=0.148	p=0.495	p=0.743	p=0.303	p=0.569	p=0.079
Mean control	-0.210	0.152	0.006	-0.077	-0.001	0.871	0.677
Low achievers	-0.542	-0.347	-0.574	-0.413	-0.528	0.809	0.501
High achievers	0.120	0.649	0.582	0.256	0.517	0.940	0.872
P-value (low = high)	0.001	0.098	0.067	0.232	0.030	0.656	0.329
Observations	907	906	903	903	901	1012	1012

Notes: Treatment effects on national exam outcomes using administrative data. Columns 1–5 show effects on normalized MSCE scores in English, Biology, Maths, and Chichewa, and the aggregate MSCE score. Column 6 shows an indicator for passing the national exam, and column 7 shows an indicator for qualifying for university admission. Panel A includes all students, Panel B restricts to students in Form 4, and Panel C restricts to students in Forms 2 and 3 at the time of the intervention. All panels show average and heterogeneous treatment effects by baseline achievement. Low (high) achievers had average baseline exam scores in English and Biology below (above) the median. All regressions include randomization strata fixed effects and control for baseline English score. Robust standard errors in parentheses. Randomization inference p-values based on 10,000 replications denoted as "p =". Anderson (2008) sharpened q-values, in brackets, correcting for 14 pre-registered hypotheses (7 outcomes × 2 samples). P-value for a test of equal coefficients between groups. * $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$.

Table 2: Higher Education and Employment Within 3 Years of Secondary School

	Higher Education		University		Paid Employment
	(1)	(2)	(3)	(4)	(5)
	Enrolled	≥1 yr	Enrolled	≥1 yr	Any
Panel A: Heterogeneity by Baseline Achievement					
Treatment	0.065**	0.056**	0.064***	0.053**	-0.002
	(0.029)	(0.027)	(0.024)	(0.023)	(0.021)
	p=0.026	p=0.043	p=0.013	p=0.024	p=0.903
Treatment × low achiever	0.097**	0.086**	0.091**	0.075**	-0.022
	(0.044)	(0.042)	(0.038)	(0.037)	(0.029)
	p=0.030	p=0.049	p=0.021	p=0.051	p=0.507
Treatment × high achiever	0.032	0.026	0.036	0.031	0.017
	(0.037)	(0.034)	(0.030)	(0.026)	(0.029)
	p=0.409	p=0.449	p=0.252	p=0.333	p=0.520
Mean control	0.080	0.072	0.048	0.044	0.056
Low achievers	0.085	0.078	0.054	0.054	0.062
High achievers	0.074	0.066	0.041	0.033	0.050
P-value (low = high)	0.256	0.267	0.262	0.326	0.344
Observations	511	511	511	511	511
Panel B: Heterogeneity by Gender					
Treatment × Female	0.146***	0.128***	0.134***	0.118***	-0.004
	(0.044)	(0.041)	(0.039)	(0.035)	(0.027)
	p=0.001	p=0.004	p=0.001	p=0.001	p=0.864
Treatment × Male	-0.002	-0.003	0.006	-0.000	0.000
	(0.037)	(0.036)	(0.030)	(0.029)	(0.030)
	p=0.955	p=0.944	p=0.839	p=0.992	p=0.999
Mean control	0.080	0.072	0.048	0.044	0.056
Female	0.054	0.045	0.027	0.018	0.036
Male	0.101	0.094	0.065	0.065	0.072
P-value (female = male)	0.012	0.018	0.012	0.011	0.923
Observations	511	511	511	511	511
Panel C: Heterogeneity by Socioeconomic Status					
Treatment × low SES	0.043	0.033	0.029	0.019	0.010
	(0.036)	(0.035)	(0.026)	(0.024)	(0.033)
	p=0.248	p=0.370	p=0.264	p=0.441	p=0.767
Treatment × high SES	0.072	0.068	0.085**	0.075**	-0.010
	(0.044)	(0.042)	(0.040)	(0.037)	(0.026)
	p=0.119	p=0.120	p=0.045	p=0.062	p=0.698
Mean control	0.080	0.072	0.048	0.044	0.056
Low SES	0.053	0.053	0.023	0.023	0.068
High SES	0.111	0.094	0.077	0.068	0.043
P-value (low SES = high SES)	0.611	0.520	0.239	0.205	0.645
Observations	511	511	511	511	511

Notes: Treatment effects on higher education and employment outcomes within three years of secondary school graduation, using survey data. Columns 1 and 3 show indicators for enrolling in any higher education institution or in a university within three years. Columns 2 and 4 show indicators for completing at least one year in higher education or in a university within three years. Column 5 is an indicator for any paid employment within three years. Panel A shows average and heterogeneous treatment effects by baseline achievement. Low (high) achievers had average baseline exam scores in English and Biology below (above) the median. Panel B shows heterogeneous effects by gender. Panel C shows heterogeneous effects by socioeconomic status (SES). High SES is an indicator variable for having both running water and electricity at home at baseline. Panels B and C extend Equation 1 to include treatment interactions with gender and SES, respectively, plus main effects. All regressions include randomization strata fixed effects and control for baseline English score. Robust standard errors in parentheses. Randomization inference p-values based on 10,000 replications denoted as "p =". P-value for a test of equal coefficients between groups. * $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$.

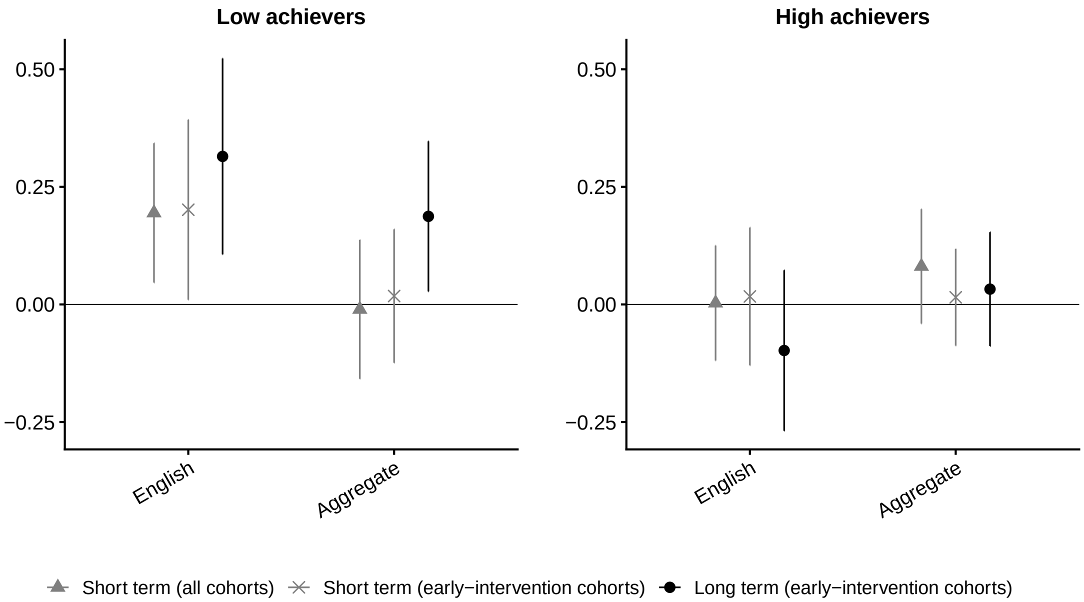
Table 3: National Exam Scores by Socioeconomic Status and Gender

	(1) English	(2) Biology	(3) Maths	(4) Chichewa	(5) Aggregate	(6) Pass	(7) Qualified for University
			Panel A: Heterogeneity by Gender				
Treatment × Female	0.059	0.092	0.065	-0.009	0.100*	0.022	0.082**
	(0.063)	(0.061)	(0.076)	(0.064)	(0.057)	(0.018)	(0.033)
	p=0.314	p=0.111	p=0.372	p=0.882	p=0.081	p=0.262	p=0.013
Treatment × Male	0.104*	0.167***	0.119	0.046	0.177***	0.015	0.065*
	(0.054)	(0.063)	(0.079)	(0.061)	(0.061)	(0.028)	(0.035)
	p=0.061	p=0.008	p=0.108	p=0.455	p=0.003	p=0.609	p=0.066
Mean control	0.000	0.000	0.000	-0.000	0.000	0.892	0.701
Female	0.171	0.037	-0.034	0.113	0.150	0.951	0.767
Male	-0.139	-0.030	0.028	-0.096	-0.121	0.849	0.653
P-value (female = male)	0.583	0.396	0.627	0.533	0.352	0.832	0.713
Observations	1381	1384	1377	1354	1378	1508	1508
			Panel B: Heterogeneity by Socioeconomic Status				
Treatment × low SES	0.068	0.120*	0.126	-0.021	0.148**	-0.011	0.086**
	(0.058)	(0.063)	(0.083)	(0.062)	(0.061)	(0.027)	(0.036)
	p=0.218	p=0.055	p=0.095	p=0.720	p=0.011	p=0.670	p=0.014
Treatment × high SES	0.090	0.133**	0.058	0.077	0.131**	0.050**	0.060*
	(0.060)	(0.064)	(0.075)	(0.064)	(0.060)	(0.022)	(0.034)
	p=0.133	p=0.034	p=0.432	p=0.242	p=0.028	p=0.049	p=0.076
Mean control	-0.000	0.001	-0.000	-0.001	0.001	0.892	0.701
Low SES	-0.218	-0.075	0.020	-0.000	-0.088	0.897	0.691
High SES	0.246	0.088	-0.023	-0.001	0.101	0.887	0.713
P-value (low SES = high SES)	0.798	0.882	0.546	0.283	0.839	0.083	0.598
Observations	1379	1382	1375	1352	1376	1506	1506

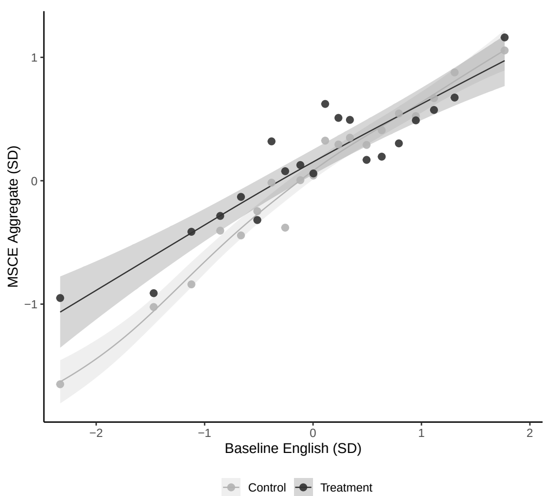
Notes: Heterogeneous treatment effects on national exam outcomes using administrative data. Columns 1–5 show effects on normalized MSCE scores in English, Biology, Maths, and Chichewa, and the aggregate MSCE score. Column 6 shows an indicator for passing the national exam, and column 7 shows an indicator for qualifying for university admission. Panel A shows heterogeneous effects by gender. Panel B shows heterogeneous effects by socioeconomic status (SES). High SES is an indicator variable for having both running water and electricity at home at baseline. Both panels extend Equation 1 to include treatment interactions with gender and SES, respectively, plus main effects. All regressions include randomization strata fixed effects and control for baseline English score. Robust standard errors in parentheses. Randomization inference p-values based on 10,000 replications denoted as "p =". P-value for a test of equal coefficients between groups. * $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$.

Figure 1: Treatment Effects on Exam Scores

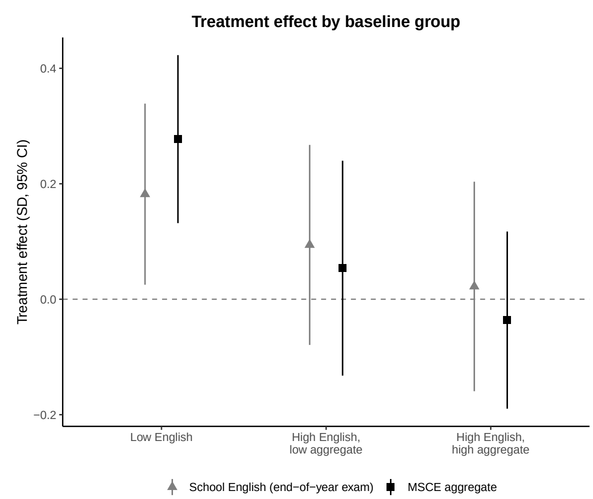
A: Treatment Effects Over Time



B: MSCE Aggregate vs. Baseline English



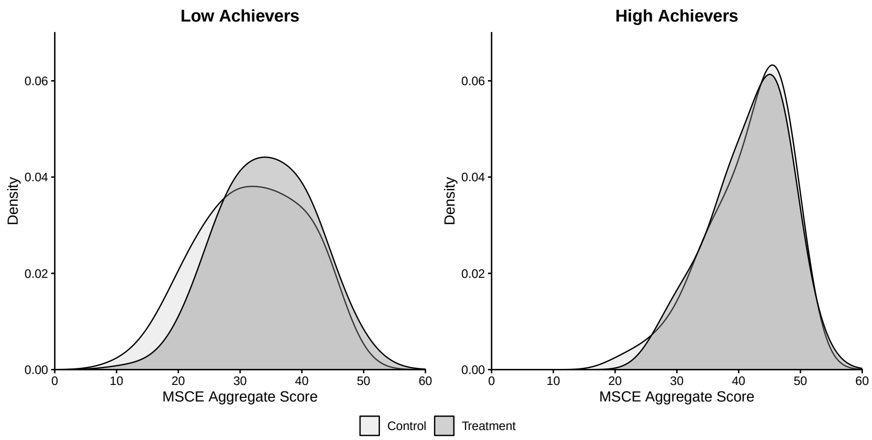
C: Treatment Effects by Baseline Group



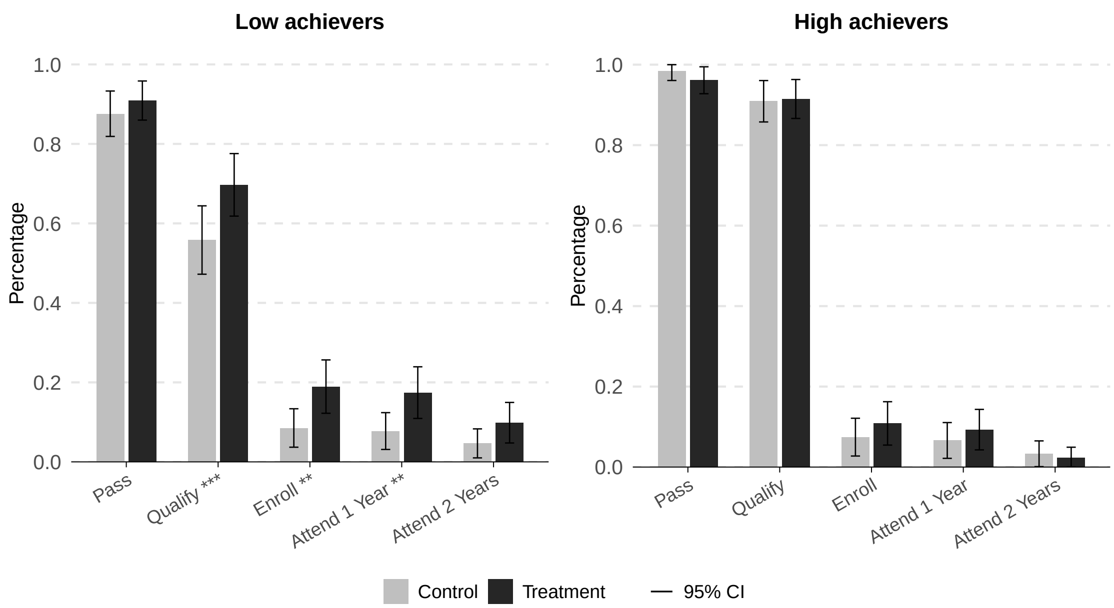
Notes: Panel A shows regression estimates for end-of-year school exam scores and national final exam (MSCE) scores. For comparability, we use the specification in Derksen et al. (2022): with strata fixed effects and control for the corresponding baseline score (and missingness indicator). Baseline and short-term aggregate scores are constructed as the average of top 6 raw subject scores (or missing if fewer than 6 scores), using pre-intervention and end-of-year school exam scores, respectively. "Early-intervention cohorts" received the intervention before their final year (students in Forms 2 and 3 at the time of the intervention). Low (high) achievers had average baseline exam scores in English and Biology below (above) the median. Panel B is a binscatter of MSCE aggregate scores against baseline English by treatment status, using 20 pooled quantile-based bins (same bin cutoffs for both groups). Each dot shows the within-group mean of MSCE aggregate in the bin. Fitted curves are cubic-spline GAM smoothers; shaded bands are pointwise 95% confidence intervals. Panel C shows treatment effects on MSCE aggregate and end-of-year English scores by three baseline groups: (i) below-median English (or missing), (ii) above-median English and below-median aggregate (or missing), and (iii) above-median English and above-median aggregate. Baseline aggregate is constructed as the average of top 6 raw subject scores (or missing if fewer than 6 scores). Estimates are from OLS regressions of each outcome on the three group dummies and their interactions with treatment, with strata fixed effects. All panels: For comparability with non-standardized school exams, outcomes are normalized within school and form. Heteroskedasticity-robust 95% confidence intervals.

Figure 2: Distributions of National Exam Scores and Pipeline Towards Higher Education

A: Aggregate Score



B: Graduation and Higher Education (Within 3 Years)



Notes: Panel A shows densities of the aggregate MSCE score. Left panel shows low achievers and right panel shows high achievers, by treatment status. Scores reversed so that higher values indicate better performance (60 minus raw score). Low (high) achievers had average baseline exam scores in English and Biology below (above) the median. Distributions of scores by subject are shown in Figure A4. Panel B shows mean rates of passing the national exam, qualifying for university, and higher education within three years of secondary school graduation. * $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$.

ONLINE APPENDIX

A Additional Tables and Figures

A.1 Appendix Tables

Table A1: Balance

	(1) Treatment	(2) Control (subsample)	(3) P-values	(4) Control (full)	(5) P-values
Average exam score in English	57.188	57.626	0.694	57.429	0.780
	(13.187)	(13.533)		(13.005)	
Average exam score in Biology	53.810	53.267	0.709	53.544	0.816
	(17.196)	(17.640)		(17.985)	
Average exam score in Maths	44.892	43.535	0.476	43.899	0.501
	(21.904)	(23.077)		(22.622)	
Average exam score in Chichewa	61.791	61.712	0.947	62.007	0.819
	(14.438)	(13.782)		(13.942)	
Age	15.973	16.060	0.577	16.033	0.635
	(1.971)	(1.845)		(1.869)	
Female	0.452	0.433	0.641	0.423	0.361
	(0.499)	(0.496)		(0.494)	
District of origin	0.605	0.574	0.444	0.575	0.348
	(0.490)	(0.495)		(0.495)	
Mother's education	0.746	0.698	0.224	0.718	0.358
	(0.436)	(0.460)		(0.450)	
Father's education	0.849	0.852	0.918	0.856	0.775
	(0.359)	(0.356)		(0.351)	
Household has electricity	0.611	0.557	0.179	0.576	0.262
	(0.488)	(0.498)		(0.494)	
Household has a mobile phone	0.870	0.849	0.451	0.866	0.852
	(0.336)	(0.359)		(0.340)	
Observations	301	298		1,207	

Notes: From Derksen et al. (2022). Balance table between treatment and control groups. Columns 1, 2, and 4 present means for the treatment, control with supplementary surveys, and full control group, respectively. Standard deviations are in parentheses. P-values in Columns 3 and 5 are from regressions of each characteristic on an indicator for treatment, with strata fixed effects.

Table A2: Attrition

	(1) Treatment	(2) Control	(3) P-value
	Panel A. Exam Score is Missing		
English	0.073	0.087	0.422
	(0.261)	(0.282)	
Biology	0.066	0.086	0.219
	(0.249)	(0.281)	
Maths	0.070	0.091	0.204
	(0.255)	(0.288)	
Chichewa	0.086	0.106	0.284
	(0.281)	(0.308)	
Aggregate	0.080	0.088	0.662
	(0.271)	(0.283)	
Pass	0.000	0.000	
University Qualification	0.000	0.000	
Number of students	301	1207	
	Panel B. Survey Data is Missing		
Endline Survey	0.083	0.084	0.933
	(0.276)	(0.278)	
Long-term Survey	0.133	0.161	0.362
	(0.340)	(0.368)	
Number of students	301	298	

Notes: Attrition rates in the treatment and control groups. In Panel A, the outcome is an indicator variable that equals one if MSCE data for a given school subject or aggregate score is missing. The control group includes all control students. Pass and university qualification are set to zero for all students who did not pass or qualify with their cohort, so there is no missing data for these outcomes. In Panel B, the outcome is an indicator that equals one if the individual is missing in the survey data. The control group includes only students with supplementary surveys. Columns 1 and 2 show the mean attrition rates for treatment and control groups, respectively. Standard deviations in parentheses. P-values in Column 3 are from regressions of the outcome on a treatment indicator, with strata fixed effects.

Table A3: National Exam Pass and Credit by Subject

	English		Biology		Maths		Chichewa	
	(1) Pass	(2) Credit	(3) Pass	(4) Credit	(5) Pass	(6) Credit	(7) Pass	(8) Credit
Treatment	0.005	0.043**	0.019	0.060**	0.034	0.050*	0.016	0.028
	(0.017)	(0.022)	(0.016)	(0.023)	(0.021)	(0.028)	(0.018)	(0.021)
Treatment × low achiever	0.017	0.091***	0.038	0.119***	0.057	0.116***	0.033	0.058*
	(0.026)	(0.033)	(0.025)	(0.037)	(0.035)	(0.042)	(0.027)	(0.032)
Treatment × high achiever	-0.009	-0.010	-0.004	-0.007	0.008	-0.025	-0.004	-0.006
	(0.021)	(0.027)	(0.020)	(0.027)	(0.021)	(0.035)	(0.023)	(0.026)
Mean control	0.909	0.800	0.902	0.744	0.824	0.606	0.891	0.828
Low achievers	0.871	0.706	0.857	0.592	0.721	0.389	0.849	0.750
High achievers	0.951	0.905	0.953	0.914	0.940	0.847	0.939	0.914
P-value (low = high)	0.430	0.018	0.195	0.006	0.234	0.011	0.304	0.120
Observations	1508	1508	1508	1508	1508	1508	1508	1508

Notes: Treatment effects on likelihood to pass or get credit on national exam in English, Biology, Maths, and Chichewa. The outcome in odd columns is the likelihood to pass the exam in the given school subject (i.e. a raw score of 8 or better). The outcome in even columns is the likelihood to get a credit in the given school subject (i.e. a raw score of 6 or better). Low (high) achievers had average baseline exam scores in English and Biology below (above) the median. All regressions include randomization strata fixed effects and control for baseline English score. Robust standard errors in parentheses. P-value for a test of equal coefficients between groups. * $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$.

Table A4: Higher Education and Employment by Cohort

	Higher Education		University		Paid Employment
	(1) Enrolled	(2) ≥1 yr	(3) Enrolled	(4) ≥1 yr	(5) Any
Panel A: Same-year intervention (Form 4)					
Treatment	0.135**	0.095*	0.114**	0.066*	0.018
	(0.054)	(0.048)	(0.045)	(0.039)	(0.045)
Treatment × low achiever	0.175**	0.142**	0.128*	0.079	0.020
	(0.077)	(0.069)	(0.068)	(0.060)	(0.062)
Treatment × high achiever	0.092	0.044	0.100*	0.052	0.015
	(0.076)	(0.069)	(0.059)	(0.048)	(0.065)
Mean control	0.071	0.060	0.036	0.036	0.071
Low achievers	0.067	0.044	0.044	0.044	0.067
High achievers	0.077	0.077	0.026	0.026	0.077
P-value (low = high)	0.449	0.318	0.756	0.724	0.949
Observations	167	167	167	167	167
Panel B: Earlier intervention (Forms 2 and 3)					
Treatment	0.029	0.037	0.038	0.045	-0.011
	(0.033)	(0.033)	(0.029)	(0.028)	(0.021)
Treatment × low achiever	0.054	0.054	0.068	0.068	-0.040
	(0.053)	(0.053)	(0.046)	(0.046)	(0.030)
Treatment × high achiever	0.005	0.019	0.009	0.023	0.017
	(0.041)	(0.038)	(0.034)	(0.031)	(0.029)
Mean control	0.084	0.078	0.054	0.048	0.048
Low achievers	0.095	0.095	0.060	0.060	0.060
High achievers	0.073	0.061	0.049	0.037	0.037
P-value (low = high)	0.461	0.592	0.305	0.424	0.174
Observations	344	344	344	344	344

Notes: Treatment effects on higher education and employment outcomes within three years of secondary school graduation, using survey data. Columns 1 and 3 show indicators for enrolling in any higher education institution or in a university within three years. Columns 2 and 4 show indicators for completing at least one year in higher education or in a university within three years. Column 5 is an indicator for any paid employment within three years. Panel A restricts to students in Form 4 and Panel B restricts to students in Forms 2 and 3 at the time of the intervention. All panels show average and heterogeneous treatment effects by baseline achievement. Low (high) achievers had average baseline exam scores in English and Biology below (above) the median. All regressions include randomization strata fixed effects and control for baseline English score. Robust standard errors in parentheses. P-value for a test of equal coefficients between groups. We also test for differences in treatment effects between the same-year (Form 4) and earlier (Forms 2–3) intervention cohorts, and only the average treatment effect in column (1) differs significantly at the 10 percent level. * $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$.

Table A5: Higher Education by Institution Type

	Accredited		Public		Private	
	(1) Enrolled	(2) $\geq$1 yr	(3) Enrolled	(4) $\geq$1 yr	(5) Enrolled	(6) $\geq$1 yr
Treatment	0.050**	0.043**	0.020	0.017	0.036*	0.028
	(0.022)	(0.020)	(0.016)	(0.014)	(0.018)	(0.018)
Treatment $\times$ low achiever	0.071**	0.063*	0.026	0.026	0.065**	0.049
	(0.034)	(0.034)	(0.021)	(0.021)	(0.033)	(0.031)
Treatment $\times$ high achiever	0.029	0.023	0.014	0.008	0.006	0.006
	(0.027)	(0.023)	(0.023)	(0.018)	(0.016)	(0.016)
Mean control	0.040	0.036	0.020	0.016	0.028	0.028
Low achievers	0.047	0.047	0.016	0.016	0.039	0.039
High achievers	0.033	0.025	0.025	0.017	0.017	0.017
P-value (low = high)	0.330	0.318	0.710	0.521	0.106	0.221
Observations	511	511	511	511	511	511

Notes: Treatment effects on higher education outcomes by institution type within three years of secondary school graduation, using survey data. Columns 1 and 2 show enrollment and an indicator for completing at least one year at accredited universities. Columns 3 and 4 show the same for public universities, and columns 5 and 6 for private universities. *Accredited* includes accredited public and private universities as classified by the Malawian National Council for Higher Education (NCHE). *Public (private)* includes only public (private) universities as determined by NCHE. All outcomes are measured within three years of secondary school graduation. Average and heterogeneous treatment effects by baseline achievement are shown. Low (high) achievers had average baseline exam scores in English and Biology below (above) the median. All regressions include randomization strata fixed effects and control for baseline English score. Robust standard errors in parentheses. P-value for a test of equal coefficients between groups. * $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$.

Table A6: National Exam Scores by Cohort and Gender

	(1) English	(2) Biology	(3) Maths	(4) Chichewa	(5) Aggregate	(6) Pass	(7) Qualified for University
	Panel A: Same-year intervention (Form 4)						
Treatment × Female	-0.045	0.015	0.027	-0.094	-0.062	-0.015	-0.003
	(0.108)	(0.101)	(0.129)	(0.131)	(0.093)	(0.032)	(0.064)
Treatment × Male	0.136*	0.036	0.244**	0.031	0.127	0.016	0.071
	(0.080)	(0.090)	(0.109)	(0.093)	(0.079)	(0.041)	(0.060)
Mean control	0.398	-0.287	-0.012	0.152	0.002	0.935	0.749
Female	0.517	-0.330	0.070	0.291	0.239	0.975	0.790
Male	0.317	-0.257	-0.067	0.049	-0.158	0.909	0.722
P-value (female = male)	0.178	0.874	0.200	0.441	0.122	0.556	0.401
Observations	474	478	474	451	477	496	496
	Panel B: Earlier intervention (Forms 2 and 3)						
Treatment × Female	0.114	0.134*	0.083	0.032	0.180**	0.040*	0.129***
	(0.078)	(0.075)	(0.095)	(0.072)	(0.072)	(0.022)	(0.036)
Treatment × Male	0.087	0.238***	0.050	0.052	0.199**	0.014	0.063
	(0.071)	(0.084)	(0.106)	(0.078)	(0.082)	(0.036)	(0.044)
Mean control	-0.210	0.152	0.006	-0.077	-0.001	0.871	0.677
Female	0.014	0.204	-0.081	0.032	0.109	0.941	0.756
Male	-0.412	0.106	0.085	-0.175	-0.099	0.818	0.616
P-value (female = male)	0.796	0.355	0.815	0.846	0.856	0.547	0.248
Observations	907	906	903	903	901	1012	1012

Notes: Heterogeneous treatment effects by gender on national exam outcomes using administrative data. Columns 1–5 show effects on normalized MSCE scores in English, Biology, Maths, and Chichewa, and the aggregate MSCE score. Column 6 shows an indicator for passing the national exam, and column 7 shows an indicator for qualifying for university admission. Panel A restricts to students in Form 4 and Panel B restricts to students in Forms 2 and 3 at the time of the intervention. Both panels extend Equation 1 to include treatment interactions with gender plus main effect. All regressions include randomization strata fixed effects and control for baseline English score. Robust standard errors in parentheses. P-value for a test of equal coefficients between groups. * $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$.

Table A7: National Exam Scores by Cohort and Socioeconomic Status

	(1) English	(2) Biology	(3) Maths	(4) Chichewa	(5) Aggregate	(6) Pass	(7) Qualified for University
	Panel A: Same-year intervention (Form 4)						
Treatment × low SES	0.049	0.076	0.289**	-0.057	0.131*	-0.040	0.050
	(0.102)	(0.087)	(0.119)	(0.113)	(0.073)	(0.049)	(0.062)
Treatment × high SES	0.050	-0.063	-0.012	0.020	-0.052	0.055***	0.007
	(0.092)	(0.107)	(0.120)	(0.112)	(0.097)	(0.019)	(0.066)
Mean control	0.399	-0.285	-0.012	0.152	0.004	0.934	0.750
Low SES	0.229	-0.312	0.001	0.175	-0.056	0.933	0.737
High SES	0.587	-0.254	-0.027	0.122	0.071	0.936	0.765
P-value (low SES = high SES)	0.995	0.317	0.076	0.632	0.137	0.075	0.640
Observations	472	476	472	449	475	494	494
	Panel B: Earlier intervention (Forms 2 and 3)						
Treatment × low SES	0.081	0.148*	0.040	0.002	0.168**	0.005	0.109**
	(0.070)	(0.084)	(0.110)	(0.075)	(0.085)	(0.033)	(0.044)
Treatment × high SES	0.099	0.221***	0.091	0.099	0.212***	0.049	0.082**
	(0.078)	(0.078)	(0.095)	(0.078)	(0.074)	(0.031)	(0.039)
Mean control	-0.210	0.152	0.006	-0.077	-0.001	0.871	0.677
Low SES	-0.449	0.049	0.030	-0.091	-0.105	0.879	0.668
High SES	0.063	0.270	-0.021	-0.060	0.118	0.862	0.688
P-value (low SES = high SES)	0.865	0.528	0.728	0.374	0.700	0.333	0.650
Observations	907	906	903	903	901	1012	1012

Notes: Heterogeneous treatment effects by socioeconomic status (SES) on national exam outcomes using administrative data. Columns 1–5 show effects on normalized MSCE scores in English, Biology, Maths, and Chichewa, and the aggregate MSCE score. Column 6 shows an indicator for passing the national exam, and column 7 shows an indicator for qualifying for university admission. High SES is an indicator variable for having both running water and electricity at home at baseline. Panel A restricts to students in Form 4 and Panel B restricts to students in Forms 2 and 3 at the time of the intervention. Both panels extend Equation 1 to include treatment interactions with high SES plus main effect. All regressions include randomization strata fixed effects and control for baseline English score. Robust standard errors in parentheses. P-value for a test of equal coefficients between groups. * $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$.

Table A8: Higher Education and Employment by Cohort and Gender

	Higher Education		University		Paid Employment
	(1) Enrolled	(2) ≥1 yr	(3) Enrolled	(4) ≥1 yr	(5) Any
	Panel A: Same-year intervention (Form 4)				
Treatment × Male	0.024	0.024	0.036	0.016	0.036
	(0.067)	(0.060)	(0.051)	(0.045)	(0.069)
Treatment × Female	0.267***	0.180**	0.207***	0.126*	-0.004
	(0.086)	(0.078)	(0.078)	(0.067)	(0.058)
Mean control	0.071	0.060	0.036	0.036	0.071
Female	0.050	0.050	0.025	0.025	0.050
Male	0.091	0.068	0.045	0.045	0.091
P-value (female = male)	0.032	0.118	0.073	0.180	0.657
Observations	167	167	167	167	167
	Panel B: Earlier intervention (Forms 2 and 3)				
Treatment × Male	-0.013	-0.014	-0.006	-0.006	-0.018
	(0.045)	(0.045)	(0.037)	(0.037)	(0.031)
Treatment × Female	0.081	0.098**	0.092**	0.109***	-0.001
	(0.050)	(0.047)	(0.044)	(0.040)	(0.028)
Mean control	0.084	0.078	0.054	0.048	0.048
Female	0.056	0.042	0.028	0.014	0.028
Male	0.105	0.105	0.074	0.074	0.063
P-value (female = male)	0.164	0.088	0.093	0.038	0.699
Observations	344	344	344	344	344

Notes: Heterogeneous treatment effects on higher education and employment outcomes by gender within three years of secondary school graduation, using survey data. Columns 1 and 3 show indicators for enrolling in any higher education institution or in a university within three years. Columns 2 and 4 show indicators for completing at least one year in higher education or in a university within three years. Column 5 is an indicator for any paid employment within three years. Panel A restricts to students in Form 4 and Panel B restricts to students in Forms 2 and 3 at the time of the intervention. Both panels extend Equation 1 to include treatment interactions with gender plus main effect. All regressions include randomization strata fixed effects and control for baseline English score. Robust standard errors in parentheses. P-value for a test of equal coefficients between groups. * $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$.

Table A9: Higher Education and Employment by Cohort and Socioeconomic Status

	Higher Education		University		Paid Employment
	(1) Enrolled	(2) $\geq$1 yr	(3) Enrolled	(4) $\geq$1 yr	(5) Any
Panel A: Same-year intervention (Form 4)					
Treatment × low SES	0.124**	0.095*	0.091**	0.062*	0.030
	(0.062)	(0.057)	(0.046)	(0.037)	(0.064)
Treatment × high SES	0.135	0.086	0.134	0.065	0.004
	(0.095)	(0.086)	(0.081)	(0.070)	(0.069)
Mean control	0.071	0.060	0.036	0.036	0.071
Low SES	0.022	0.022	0.000	0.000	0.065
High SES	0.132	0.105	0.079	0.079	0.079
P-value (low SES = high SES)	0.925	0.934	0.652	0.967	0.782
Observations	167	167	167	167	167
Panel B: Earlier intervention (Forms 2 and 3)					
Treatment × low SES	-0.002	-0.001	-0.004	-0.003	-0.002
	(0.045)	(0.045)	(0.031)	(0.031)	(0.037)
Treatment × high SES	0.047	0.060	0.063	0.077*	-0.014
	(0.049)	(0.047)	(0.046)	(0.044)	(0.022)
Mean control	0.084	0.078	0.054	0.048	0.048
Low SES	0.069	0.069	0.034	0.034	0.069
High SES	0.101	0.089	0.076	0.063	0.025
P-value (low SES = high SES)	0.463	0.349	0.226	0.143	0.800
Observations	344	344	344	344	344

Notes: Heterogeneous treatment effects on higher education and employment outcomes by socioeconomic status (SES) within three years of secondary school graduation, using survey data. Columns 1 and 3 show indicators for enrolling in any higher education institution or in a university within three years. Columns 2 and 4 show indicators for completing at least one year in higher education or in a university within three years. Column 5 is an indicator for any paid employment within three years. High SES is an indicator variable for having both running water and electricity at home at baseline. Panel A restricts to students in Form 4 and Panel B restricts to students in Forms 2 and 3 at the time of the intervention. Both panels extend Equation 1 to include treatment interactions with high SES plus main effect. All regressions include randomization strata fixed effects and control for baseline English score. Robust standard errors in parentheses. P-value for a test of equal coefficients between groups. * $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$.

Table A10: Lee (2009) Bounds on National Exam Treatment Effects

	(1) English	(2) Biology	(3) Maths	(4) Chichewa	(5) Aggregate	(6) Pass	(7) Qualified for University
Panel A: All Students							
Treatment lower bound	0.033	0.053	0.013	-0.051	0.104	0.018	0.067**
	(0.071)	(0.072)	(0.074)	(0.077)	(0.068)	(0.019)	(0.028)
Treatment upper bound	0.096	0.134*	0.093	0.044	0.141*	0.018	0.067**
	(0.077)	(0.074)	(0.073)	(0.076)	(0.072)	(0.019)	(0.028)
Low achiever lower	0.119	0.103	0.120	-0.038	0.210**	0.041	0.124***
	(0.106)	(0.100)	(0.099)	(0.101)	(0.093)	(0.029)	(0.043)
Low achiever upper	0.272**	0.263***	0.261***	0.143	0.292***	0.041	0.124***
	(0.109)	(0.101)	(0.093)	(0.099)	(0.094)	(0.029)	(0.043)
High achiever lower	-0.079	0.008	-0.071	-0.050	-0.007	-0.008	0.004
	(0.096)	(0.088)	(0.093)	(0.107)	(0.085)	(0.022)	(0.028)
High achiever upper	-0.052	0.016	-0.069	-0.041	0.010	-0.008	0.004
	(0.088)	(0.076)	(0.084)	(0.099)	(0.073)	(0.022)	(0.028)
Observations	1508	1508	1508	1508	1508	1508	1508
Panel B: Same-year intervention (Form 4)							
Treatment lower bound	0.019	-0.051	0.098	-0.091	-0.011	0.004	0.027
	(0.131)	(0.103)	(0.105)	(0.120)	(0.118)	(0.027)	(0.048)
Treatment upper bound	0.052	-0.026	0.111	-0.030	0.004	0.004	0.027
	(0.116)	(0.107)	(0.109)	(0.120)	(0.107)	(0.027)	(0.048)
Low achiever lower	0.043	0.032	0.235*	-0.107	0.099	0.061*	0.155**
	(0.169)	(0.123)	(0.139)	(0.125)	(0.134)	(0.034)	(0.071)
Low achiever upper	0.185	0.196	0.385***	0.115	0.199	0.061*	0.155**
	(0.164)	(0.126)	(0.137)	(0.119)	(0.132)	(0.034)	(0.071)
High achiever lower	-0.089	-0.257**	-0.166	-0.172	-0.211	-0.060	-0.120**
	(0.138)	(0.130)	(0.138)	(0.163)	(0.133)	(0.044)	(0.059)
High achiever upper	0.061	-0.140	-0.056	-0.030	-0.092	-0.060	-0.120**
	(0.128)	(0.126)	(0.131)	(0.167)	(0.128)	(0.044)	(0.059)
Observations	496	496	496	496	496	496	496
Panel C: Earlier intervention (Forms 2 and 3)							
Treatment lower bound	0.027	0.105	-0.031	-0.031	0.155*	0.025	0.086**
	(0.086)	(0.089)	(0.097)	(0.099)	(0.087)	(0.024)	(0.034)
Treatment upper bound	0.137*	0.215**	0.085	0.083	0.219**	0.025	0.086**
	(0.081)	(0.097)	(0.096)	(0.096)	(0.092)	(0.024)	(0.034)
Low achiever lower	0.164	0.139	0.056	-0.001	0.273**	0.033	0.110**
	(0.133)	(0.141)	(0.134)	(0.134)	(0.132)	(0.040)	(0.053)
Low achiever upper	0.321**	0.296**	0.197	0.153	0.342***	0.033	0.110**
	(0.133)	(0.138)	(0.120)	(0.131)	(0.132)	(0.040)	(0.053)
High achiever lower	-0.107	0.083	-0.086	-0.046	0.052	0.018	0.064**
	(0.104)	(0.085)	(0.109)	(0.123)	(0.087)	(0.024)	(0.030)
High achiever upper	-0.051	0.139	-0.021	0.024	0.098	0.018	0.064**
	(0.106)	(0.098)	(0.113)	(0.128)	(0.091)	(0.024)	(0.030)
Observations	1012	1012	1012	1012	1012	1012	1012

Notes: Lee (2009) bounds on treatment effects. Columns 1–5 use normalized MSCE subject and aggregate scores. Bounds account for differential attrition under the monotonicity assumption that treatment weakly increases the probability of having an observed score. Pass and university qualification (Columns 6–7) have no attrition. Low (high) achievers had average baseline exam scores in English and Biology below (above) the median. Panel A includes all students, Panel B restricts to students in Form 4, and Panel C restricts to students in Forms 2 and 3 at the time of the intervention. Analytic standard errors, in parentheses, follow Tauchmann (2014). * $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$.

Table A11: Lee (2009) Bounds on National Exam Treatment Effects by Gender and Socioeconomic Status

	(1) English	(2) Biology	(3) Maths	(4) Chichewa	(5) Aggregate	(6) Pass	(7) Qualified for University
Panel A: Heterogeneity by Gender							
Female lower	0.034	0.015	0.012	-0.045	0.069	0.012	0.064*
	(0.098)	(0.100)	(0.095)	(0.099)	(0.105)	(0.019)	(0.037)
Female upper	0.097	0.067	0.038	0.020	0.086	0.012	0.064*
	(0.104)	(0.103)	(0.095)	(0.105)	(0.093)	(0.019)	(0.037)
Male lower	0.023	0.093	0.029	-0.053	0.117	0.017	0.062
	(0.100)	(0.102)	(0.109)	(0.117)	(0.095)	(0.030)	(0.040)
Male upper	0.060	0.175	0.137	0.032	0.175*	0.017	0.062
	(0.110)	(0.108)	(0.110)	(0.112)	(0.100)	(0.030)	(0.040)
Observations	1508	1508	1508	1508	1508	1508	1508
Panel B: Heterogeneity by Socioeconomic Status							
Low SES lower	-0.053	-0.042	-0.035	-0.173	0.064	-0.007	0.068*
	(0.106)	(0.098)	(0.098)	(0.111)	(0.118)	(0.029)	(0.040)
Low SES upper	0.000	0.052	0.079	-0.068	0.076	-0.007	0.068*
	(0.117)	(0.110)	(0.103)	(0.114)	(0.098)	(0.029)	(0.040)
High SES lower	0.069	0.124	0.067	0.060	0.112	0.043*	0.062
	(0.088)	(0.099)	(0.105)	(0.108)	(0.094)	(0.025)	(0.039)
High SES upper	0.122	0.191*	0.111	0.158	0.183*	0.043*	0.062
	(0.097)	(0.107)	(0.105)	(0.112)	(0.099)	(0.025)	(0.039)
Observations	1506	1506	1506	1506	1506	1506	1506

Notes: Lee (2009) bounds on treatment effects, for the subgroups in Table 3. Columns 1–5 use normalized MSCE subject and aggregate scores. Bounds account for differential attrition under the monotonicity assumption that treatment weakly increases the probability of having an observed score. Pass and university qualification (Columns 6–7) have no attrition, so lower and upper bounds coincide. Panel A reports bounds separately by gender and Panel B by socioeconomic status (SES). High SES is an indicator for having both running water and electricity at home at baseline. Analytic standard errors, in parentheses, follow Tauchmann (2014). * $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$.

Table A12: Lee (2009) Bounds on Higher Education and Employment Treatment Effects

	Higher Education		University		Paid Employment
	(1) Enrolled	(2) ≥1 yr	(3) Enrolled	(4) ≥1 yr	(5) Any
Panel A: Heterogeneity by Baseline Achievement					
Treatment lower bound	0.041	0.033	0.033	0.021	-0.030
	(0.041)	(0.041)	(0.040)	(0.039)	(0.039)
Treatment upper bound	0.074**	0.067**	0.067***	0.055**	0.003
	(0.029)	(0.028)	(0.025)	(0.023)	(0.021)
Low achiever lower	0.101*	0.093	0.086	0.071	-0.020
	(0.060)	(0.059)	(0.058)	(0.058)	(0.057)
Low achiever upper	0.105**	0.097**	0.090**	0.075**	-0.016
	(0.043)	(0.042)	(0.037)	(0.036)	(0.028)
High achiever lower	-0.025	-0.033	-0.025	-0.033	-0.041
	(0.056)	(0.055)	(0.052)	(0.029)	(0.053)
High achiever upper	0.041	0.033	0.041	0.033	0.025
	(0.038)	(0.036)	(0.031)	(0.028)	(0.031)
Panel B: Heterogeneity by Gender					
Female lower	0.151***	0.134***	0.135***	0.119***	-0.002
	(0.043)	(0.041)	(0.038)	(0.034)	(0.025)
Female upper	0.160**	0.144**	0.145**	0.128**	0.007
	(0.063)	(0.062)	(0.061)	(0.059)	(0.053)
Male lower	-0.059	-0.059	-0.060	-0.065***	-0.060
	(0.059)	(0.058)	(0.056)	(0.021)	(0.057)
Male upper	0.010	0.010	0.009	0.002	0.009
	(0.038)	(0.037)	(0.031)	(0.030)	(0.033)
Panel C: Heterogeneity by Socioeconomic Status					
Low SES lower	0.018	0.009	-0.012	-0.021	-0.023
	(0.061)	(0.061)	(0.058)	(0.058)	(0.061)
Low SES upper	0.059*	0.051	0.029	0.020	0.018
	(0.036)	(0.035)	(0.024)	(0.023)	(0.034)
High SES lower	0.064	0.060	0.077	0.064	-0.019
	(0.056)	(0.054)	(0.053)	(0.052)	(0.047)
High SES upper	0.077*	0.072*	0.089**	0.076**	-0.007
	(0.045)	(0.042)	(0.041)	(0.038)	(0.025)
Observations	599	599	599	599	599

Notes: Lee (2009) bounds on treatment effects on higher education and employment outcomes within three years of secondary school graduation, using survey data. Outcomes are defined as in Table 2. Bounds account for differential survey attrition under the monotonicity assumption that treatment weakly increases the probability of being surveyed. Panel A reports bounds for the average treatment effect and separately for low and high achievers; Panels B and C report bounds by gender and socioeconomic status (SES), mirroring the subgroups in Table 2. Low (high) achievers had average baseline exam scores in English and Biology below (above) the median. High SES is an indicator for having both running water and electricity at home at baseline. Analytic standard errors, in parentheses, follow Tauchmann (2014). * $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$.

Table A13: National Exam Raw Scores

	(1) English	(2) Biology	(3) Maths	(4) Chichewa	(5) Aggregate	(6) Pass	(7) Qualified for University
	Panel A: Same-year intervention (Form 4)						
Treatment × low achiever	0.178	0.276	0.753***	0.032	1.526**	0.061*	0.153**
	(0.170)	(0.200)	(0.274)	(0.175)	(0.773)	(0.034)	(0.066)
Sharpened q-value	[0.142]	[0.114]	[0.018]	[0.323]	[0.059]	[0.073]	[0.035]
Treatment × high achiever	-0.033	-0.306	-0.171	-0.136	-1.070	-0.059	-0.111*
	(0.141)	(0.186)	(0.273)	(0.181)	(0.757)	(0.041)	(0.056)
Mean control	6.294	5.146	4.789	6.255	36.721	0.935	0.749
Low achievers	5.625	4.075	3.515	5.668	32.184	0.901	0.575
High achievers	7.033	6.330	6.176	6.885	41.731	0.973	0.946
P-value (low = high)	0.342	0.034	0.017	0.505	0.017	0.025	0.002
Observations	474	478	474	451	477	496	496
	Panel B: Earlier intervention (Forms 2 and 3)						
Treatment × low achiever	0.442***	0.572***	0.455*	0.170	2.806***	0.035	0.118**
	(0.137)	(0.189)	(0.237)	(0.131)	(0.803)	(0.036)	(0.047)
Sharpened q-value	[0.009]	[0.011]	[0.059]	[0.119]	[0.008]	[0.149]	[0.027]
Treatment × high achiever	-0.113	0.187	-0.159	-0.035	0.622	0.015	0.063**
	(0.105)	(0.134)	(0.237)	(0.111)	(0.601)	(0.024)	(0.031)
Mean control	5.275	6.044	4.831	5.886	36.698	0.871	0.677
Low achievers	4.719	5.022	3.471	5.345	31.930	0.809	0.501
High achievers	5.828	7.061	6.181	6.421	41.387	0.940	0.872
P-value (low = high)	0.001	0.098	0.067	0.232	0.030	0.656	0.329
Observations	907	906	903	903	901	1012	1012

Notes: Treatment effects on raw national exam scores exactly as specified in the analysis plan (https://www.socialscienceregistry.org/trials/3824). Raw subject scores are reversed so that higher values indicate better performance (1–9 per subject; 6–54 for the aggregate, which is the sum of the top six). Columns 1–5 show effects on English, Biology, Maths, and Chichewa, and the aggregate MSCE score. Column 6 shows an indicator for passing the national exam, and column 7 shows an indicator for qualifying for university admission. Panel A restricts to students in Form 4, and Panel B restricts to students in Forms 2 and 3 at the time of the intervention. Low (high) achievers had average baseline exam scores in English and Biology below (above) the median. All regressions include randomization strata fixed effects and control for baseline English score. Robust standard errors in parentheses. Anderson (2008) sharpened q-values, in brackets, correcting for 14 pre-registered hypotheses (7 outcomes × 2 samples). P-value for a test of equal coefficients between groups. * $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$.

Table A14: National Exam Scores – Survey Data Sample

	(1) English	(2) Biology	(3) Maths	(4) Chichewa	(5) Aggregate	(6) Pass	(7) Qualified for University
Panel A: All Students							
Treatment	0.147**	0.176***	0.100	0.019	0.168***	0.003	0.068**
	(0.057)	(0.058)	(0.073)	(0.059)	(0.056)	(0.021)	(0.033)
Treatment × low achiever	0.263***	0.282***	0.228**	0.073	0.240***	0.026	0.124**
	(0.091)	(0.090)	(0.106)	(0.092)	(0.089)	(0.036)	(0.055)
Treatment × high achiever	0.037	0.074	-0.021	-0.033	0.099	-0.020	0.010
	(0.069)	(0.072)	(0.100)	(0.075)	(0.070)	(0.021)	(0.036)
Mean control	-0.020	-0.059	-0.005	-0.003	-0.032	0.928	0.728
Low achievers	-0.315	-0.572	-0.563	-0.364	-0.516	0.876	0.558
High achievers	0.278	0.464	0.559	0.355	0.456	0.983	0.909
P-value (low = high)	0.048	0.073	0.088	0.372	0.213	0.267	0.081
Observations	485	488	484	476	486	511	511
Panel B: Same-year intervention (Form 4)							
Treatment	0.217***	0.118	0.187*	-0.001	0.123	0.009	0.044
	(0.083)	(0.090)	(0.111)	(0.103)	(0.079)	(0.033)	(0.056)
Treatment × low achiever	0.303**	0.266**	0.298*	0.083	0.195	0.058	0.161*
	(0.130)	(0.125)	(0.160)	(0.156)	(0.121)	(0.054)	(0.090)
Treatment × high achiever	0.126	-0.043	0.068	-0.089	0.045	-0.044	-0.082
	(0.101)	(0.127)	(0.153)	(0.135)	(0.100)	(0.032)	(0.063)
Mean control	0.338	-0.376	-0.076	0.196	-0.010	0.952	0.762
Low achievers	-0.035	-0.858	-0.562	-0.177	-0.430	0.911	0.600
High achievers	0.758	0.181	0.472	0.598	0.463	1.000	0.949
P-value (low = high)	0.284	0.085	0.300	0.409	0.344	0.097	0.028
Observations	163	166	164	155	165	167	167
Panel C: Earlier intervention (Forms 2 and 3)							
Treatment	0.104	0.203***	0.053	0.025	0.189**	0.000	0.078*
	(0.074)	(0.074)	(0.094)	(0.072)	(0.074)	(0.027)	(0.040)
Treatment × low achiever	0.224*	0.281**	0.179	0.059	0.255**	0.008	0.100
	(0.119)	(0.123)	(0.138)	(0.112)	(0.119)	(0.047)	(0.069)
Treatment × high achiever	-0.004	0.133	-0.060	-0.005	0.130	-0.007	0.056
	(0.091)	(0.088)	(0.129)	(0.091)	(0.092)	(0.026)	(0.043)
Mean control	-0.210	0.112	0.034	-0.101	-0.044	0.916	0.711
Low achievers	-0.476	-0.403	-0.564	-0.463	-0.567	0.857	0.536
High achievers	0.044	0.601	0.602	0.243	0.453	0.976	0.890
P-value (low = high)	0.129	0.327	0.209	0.660	0.406	0.781	0.589
Observations	322	322	320	321	321	344	344

Notes: Treatment effects on national exam outcomes using administrative data. The sample is restricted to individuals in the long-term survey sample. Columns 1–5 show effects on normalized MSCE scores in English, Biology, Maths, and Chichewa, and the aggregate MSCE score. Column 6 shows an indicator for passing the national exam, and column 7 shows an indicator for qualifying for university admission. Panel A includes all students, Panel B restricts to students in Form 4, and Panel C restricts to students in Forms 2 and 3 at the time of the intervention. All panels show average treatment effects and heterogeneous effects by baseline achievement. Low (high) achievers had average baseline exam scores in English and Biology below (above) the median. All regressions include randomization strata fixed effects and control for baseline English score. Robust standard errors in parentheses. P-value for a test of equal coefficients between groups. * $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$.

Table A15: Spillover Analysis Using Baseline Social Network Data

	(1) English	(2) Biology	(3) Maths	(4) Chichewa	(5) Aggregate	(6) Pass	(7) Qualified for University
Treatment	0.087**	0.123***	0.082	0.023	0.136***	0.019	0.071***
	(0.042)	(0.045)	(0.055)	(0.045)	(0.042)	(0.018)	(0.025)
Treatment × treated friends	-0.015	0.031	0.072**	-0.025	0.014	-0.008	-0.000
	(0.027)	(0.030)	(0.036)	(0.027)	(0.027)	(0.011)	(0.016)
Treatment × friends	-0.002	-0.012	-0.016	0.004	-0.010	0.001	-0.001
	(0.008)	(0.008)	(0.010)	(0.008)	(0.008)	(0.003)	(0.004)
Treated friends	0.007	-0.004	-0.033**	-0.005	-0.003	0.006	0.003
	(0.011)	(0.012)	(0.015)	(0.012)	(0.011)	(0.005)	(0.007)
Friends	0.005	0.008**	0.018***	0.006*	0.010***	0.002	0.003
	(0.003)	(0.003)	(0.004)	(0.003)	(0.003)	(0.001)	(0.002)
Mean control	-0.000	-0.000	-0.000	-0.000	0.000	0.892	0.701
Observations	1381	1384	1377	1354	1378	1508	1508

Notes: Treatment effects on national exam outcomes interacted with baseline social network measures. Columns 1–5 show effects on normalized MSCE scores in English, Biology, Maths, and Chichewa, and the aggregate MSCE score. Column 6 shows an indicator for passing the national exam, and column 7 shows an indicator for qualifying for university admission. "Treated friends" is the number of treated students in the individual's baseline undirected network. "Friends" is the individual's total baseline degree (number of links across all network questions). Both network measures are demeaned (centered on their overall baseline sample means) prior to estimation, so the "Treatment" coefficient represents the treatment effect for a student with the average number of (treated) friends. All regressions include randomization strata fixed effects and control for baseline English score. Robust standard errors in parentheses. See Appendix B for details. * $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$.

Table A16: Spillover Analysis with Endogenous Network Change (Comola & Prina, 2021)

	(1) English	(2) Biology	(3) Maths	(4) Chichewa	(5) Aggregate
	Panel A: Full sample, CP dynamic PE				
Δy on Treatment	1.205	1.759	0.668	0.394	1.205*
	(0.983)	(1.208)	(1.269)	(0.724)	(0.660)
Peer effect, baseline network ($G_0 \Delta y$)	-0.134	0.312	0.290	0.324	0.416
	(0.341)	(0.427)	(0.400)	(0.278)	(0.257)
Peer effect, network change ($\Delta G\, y_1$)	0.114	0.148	0.187	0.031	0.085
	(0.076)	(0.098)	(0.123)	(0.068)	(0.066)
Observations	1,318	1,322	1,295	1,290	1,329
	Panel B: Full sample, OLS				
Δy on Treatment	1.215	2.094	0.812	0.347	1.313*
	(0.990)	(1.298)	(1.320)	(0.745)	(0.683)
Observations	1,318	1,322	1,295	1,290	1,329
	Panel C: Low achievers, OLS				
Δy on Treatment	3.643***	3.983**	3.632*	1.089	2.813***
	(1.392)	(1.966)	(1.898)	(0.983)	(0.985)
Observations	643	645	626	627	651
	Panel D: High achievers, OLS				
Δy on Treatment	-1.138	0.267	-1.847	-0.365	-0.138
	(1.104)	(1.386)	(1.657)	(1.154)	(0.841)
Observations	675	677	669	663	678

Notes: Treatment effects on the change in raw exam score $\Delta y = y_1 - y_0$, following Comola and Prina (2021). Both y_0 (baseline school exam) and y_1 (endline MSCE) are placed on a common 0–100 "percent of max" scale for comparability, so coefficients are in percentage-point units. Panel A: the direct-effect coefficient ($\widetilde{J}_{itt}$) from the Comola and Prina (2021) dynamic peer-effects 2SLS specification on the within-class network, together with the two peer-effect coefficients from the same regression: the outcome peer effect ($G_0 \Delta y$, the effect of baseline friends' average score change) and the network peer effect ($\Delta G\, y_1$, the effect of changes in friendship links weighted by endline scores). Panel B: OLS of Δy on Treatment. Panels C and D: OLS of Δy on Treatment, run separately on low and high achievers; low (high) achievers had average baseline exam scores in English and Biology below (above) the median. All regressions include randomization strata fixed effects. Cluster-robust standard errors at the classroom level in parentheses. See Appendix B for details. * $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$.

Table A17: Spillover Analysis Using National Data (All Secondary Schools, 2015–2024)

	(1) English	(2) Biology	(3) Maths	(4) Chichewa	(5) Aggregate	(6) Pass	(7) Qualified for University	(8) Num. Exam Takers
Treated school in intervention years	-0.119	-0.040	0.016	-0.159	-0.019	0.022	0.010	-2.134
	(0.101)	(0.029)	(0.114)	(0.125)	(0.022)	(0.010)	(0.014)	(11.704)
Permutation p-value	[0.191]	[0.557]	[0.490]	[0.332]	[0.265]	[0.452]	[0.826]	[0.305]
Treated student in intervention years	0.064	0.097	0.072	-0.019	0.116	0.004	0.059*	
	(0.016)	(0.028)	(0.087)	(0.043)	(0.053)	(0.009)	(0.011)	
Permutation p-value	[0.114]	[0.134]	[0.611]	[0.778]	[0.261]	[0.802]	[0.064]	
Mean control	0.000	0.000	0.000	0.000	0.000	0.539	0.187	74.6
Observations	873,560	870,327	861,185	790,333	782,907	906,816	906,816	11,887
Number of schools	1,470	1,470	1,470	1,468	1,470	1,470	1,470	1,470
Number of boarding schools	76	76	76	76	76	76	76	76
Number of years	10	10	10	10	10	10	10	10
Pre-trends F-test (p)	0.604	0.683	0.778	0.655	0.787	0.983	0.741	0.915
Post-trends F-test (p)	0.631	0.980	0.910	0.827	0.959	0.975	0.989	0.880

Notes: Treatment effects on national exam outcomes using a two-way-fixed-effect specification and all secondary schools in Malawi, in years 2015 to 2024. Columns 1–5 show effects on normalized MSCE scores in English, Biology, Maths, Chichewa, and the aggregate score, respectively. Normalization is with respect to non-intervention schools in the same year. Column 6 shows an indicator for passing the national exam, and column 7 shows an indicator for qualifying for university admission. Columns 1–7 are conditional on taking at least one exam. Column 8 shows total number of exam takers at the school-year level. "Treated school in intervention years" indicates a study school during the intervention period (2018–2020), when students from the three cohorts were taking their final national exam. "Treated student in intervention years" indicates an individually randomized treated student within a study school. All regressions include school and separate year fixed effects for boarding and non-boarding schools, and control for student gender. Robust standard errors clustered at the school level in parentheses. Permutation p-values in brackets. Significance stars and p-values for pre- and post-trend tests are based on permutation inference (10,000 permutations). See Appendix B for details. * $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$.

Table A18: Cost-Effectiveness Comparison to Benchmark Interventions

	This study	Ghana scholarships	Zambia negotiation
Cost per student	$36	$480	$60
Cost per 0.1 SD in exam scores	$26	$300	—
Years of education per $100:			
three-year follow-up	0.22	—	0.16
10+ year follow-up	—	0.28	0.38
Cost per additional HE enrollee	$554	$6,200	(insignificant)
Cost per additional HE graduate	—	—	$970
Benefit-cost ratio	10.9	3.4	7.8–16.2
MVPF	Infinite	4.1	11.1–Infinite

Notes: HE = higher education. Ghana scholarships: Duflo et al. (2025, 2026); Zambia negotiation: Ashraf et al. (2020, 2025). Exam scores are the aggregate MSCE (this study) and a PISA-style cognitive assessment administered five years into the study (Ghana, 0.16 SD). Years of education per $100: this study counts years of higher education within three years of secondary graduation; Ghana and Zambia count total years of secondary and higher education. Benefit-cost ratio and MVPF: this study counts labor market benefits only (Table A19 reports alternative scenarios); Ghana values include intergenerational child benefits, at the medium value of a statistical life; Zambia values range from the education effect alone (labor market plus intergenerational child benefits) to estimates including possible HIV reduction.

Table A19: Benefit-Cost Ratio and MVPF Under Alternative Assumptions

	Wage growth 2.9%		Wage growth 4.4%	
Earnings premium	BCR	MVPF	BCR	MVPF
Half Ghana effect (5.3%)	3.7	24	5.5	Infinite
Ghana effect (10.6%)	7.4	Infinite	10.9	Infinite
Twice Ghana effect (21.1%)	14.8	Infinite	21.8	Infinite

Notes: Each cell recomputes the benefit-cost ratio and marginal value of public funds, holding costs fixed and scaling the total benefit B, which is linear in the assumed earnings premium. Wage growth of 4.4 percent is Malawi's average real GDP growth over 2010–2019; 2.9 percent is the average over 2015–2024. The premium scenarios bound the possibility that returns for our marginal enrollees differ from the Ghana-scaled effect of 10.6 percent; the upper scenario reflects the higher observed wage premium in Malawi (Figure A6). The MVPF is infinite whenever tax revenue exceeds direct government costs. See Appendix C for details.

Table A20: Years of Higher Education Within 3 Years of Secondary School

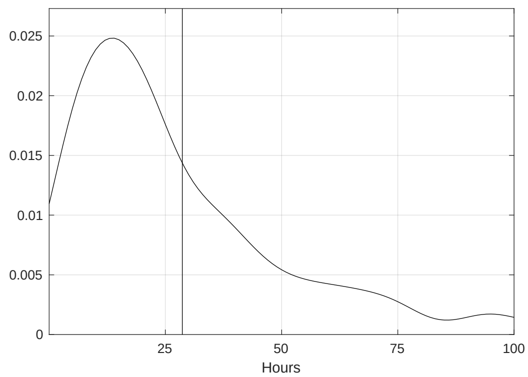
	Higher Education	University
	(1)	(2)
	Years	Years
Panel A: Heterogeneity by Baseline Achievement		
Treatment	0.075*	0.078**
	(0.043)	(0.037)
Treatment × low achiever	0.133*	0.125**
	(0.071)	(0.063)
Treatment × high achiever	0.016	0.030
	(0.049)	(0.037)
Mean control	0.112	0.068
Low achievers	0.124	0.085
High achievers	0.099	0.050
P-value (low = high)	0.176	0.193
Observations	511	511
Panel B: Heterogeneity by Gender		
Treatment × Female	0.192***	0.184***
	(0.062)	(0.053)
Treatment × Male	-0.022	-0.009
	(0.058)	(0.050)
Mean control	0.112	0.068
Female	0.063	0.018
Male	0.151	0.108
P-value (female = male)	0.012	0.008
Observations	511	511
Panel C: Heterogeneity by Socioeconomic Status		
Treatment × low SES	0.041	0.032
	(0.053)	(0.037)
Treatment × high SES	0.089	0.106*
	(0.069)	(0.061)
Mean control	0.112	0.068
Low SES	0.075	0.030
High SES	0.154	0.111
P-value (low SES = high SES)	0.590	0.296
Observations	511	511

Notes: Treatment effects on the number of years spent in any higher education institution (column 1) or in a university (column 2) within three years of secondary school graduation, using survey data. Years include ongoing enrollment spells, measured up to the survey date, and are capped at three. Panel A shows average treatment effects and heterogeneous effects by baseline achievement. Low (high) achievers had average baseline exam scores in English and Biology below (above) the median. Panel B shows heterogeneous effects by gender. Panel C shows heterogeneous effects by socioeconomic status (SES). High SES is an indicator variable for having both running water and electricity at home at baseline. Panels B and C extend Equation 1 to include treatment interactions with gender and SES, respectively, plus main effects. All regressions include randomization strata fixed effects and control for baseline English score. Robust standard errors in parentheses. P-value for a test of equal coefficients between groups. * $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$.

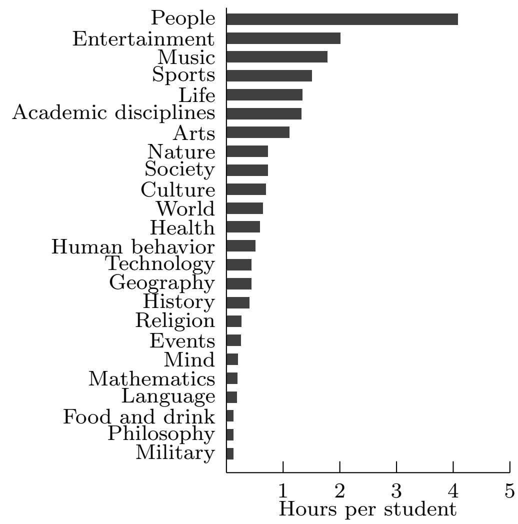
A.2 Appendix Figures

Appendix Figure A1: Browsing Hours: Distribution and Topic Popularity

Panel A. Distribution of browsing hours

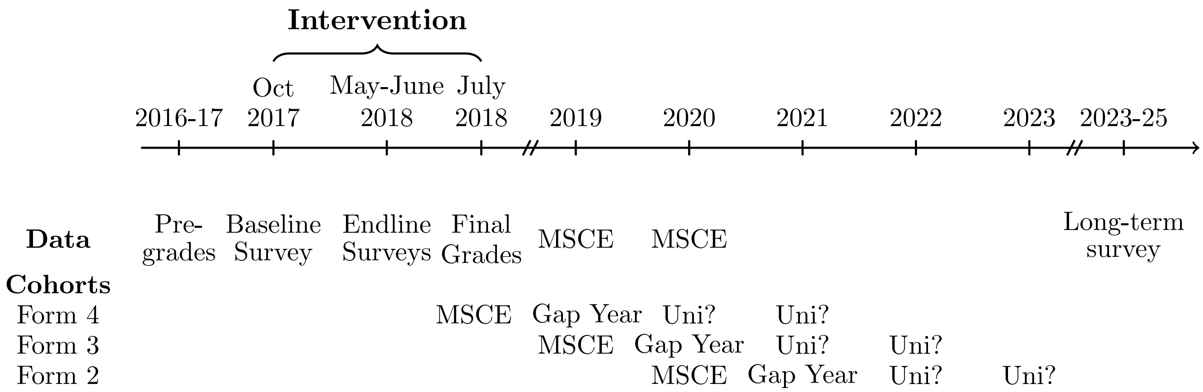


Panel B. Topic popularity



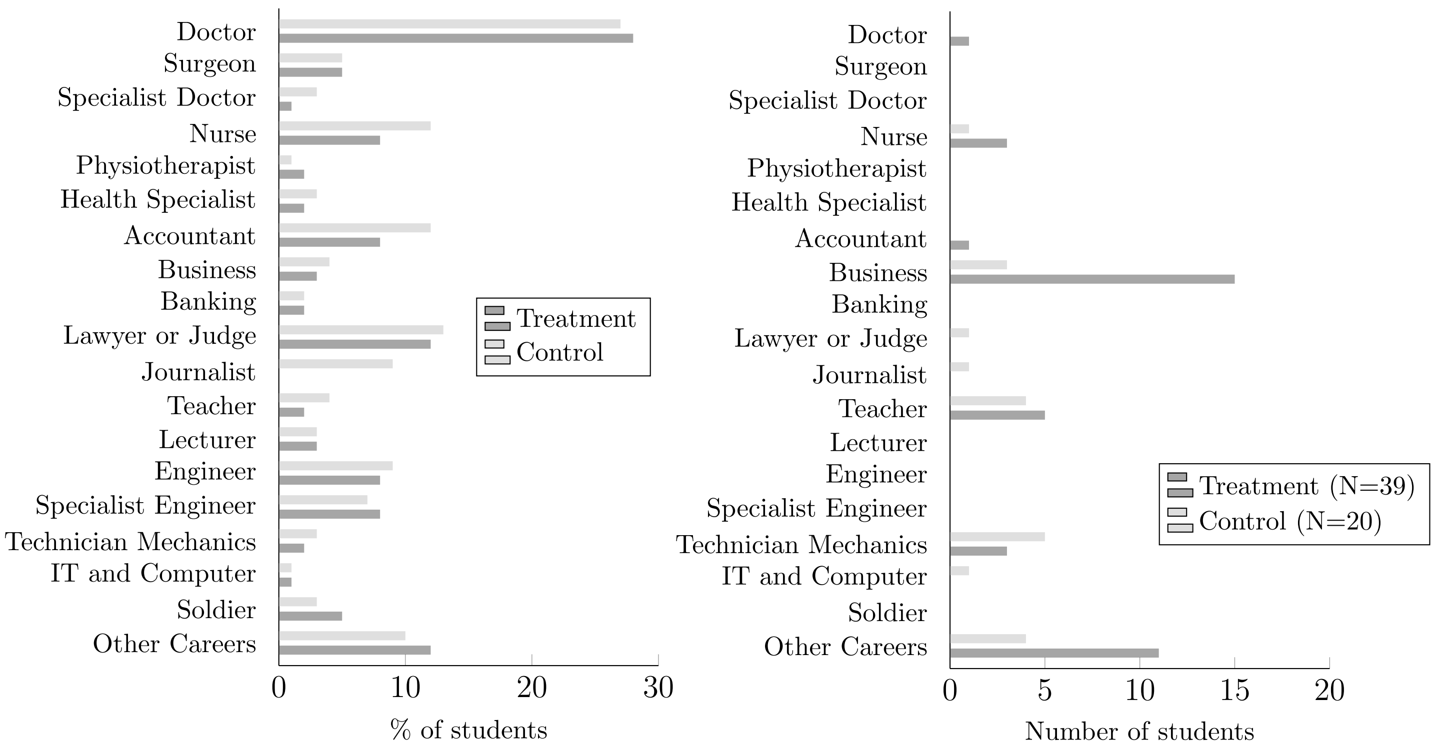
Notes: From Derksen et al. (2022). Panel A: Density of browsing hours, treatment students only, aggregated over one academic year. The digital library was open for 20–22 weeks, from November 2017 to June 2018, excluding Christmas and Easter vacations. Vertical line is the average hours spent browsing. Average is 28.6. Panel B: Browsing hours per topic, per student, aggregated over one academic year. See Derksen et al. (2022) for details on topic classification. The topics Business, Concepts, Crime, Economy, Education, Energy, Government, Humanities, Knowledge, Law, Objects, Organizations, Politics, Science, and Universe are excluded from the figure and are less than 0.12 hours.

Appendix Figure A2: Project Timeline



Notes: Timeline showing intervention, data collection, and student expected schooling trajectories by cohort. The intervention, including data collection, ran from October 2017 to July 2018. The digital library was open for 20–22 weeks, from November 2017 to June 2018, excluding Christmas and Easter vacations. We collected data from the three cohorts who participated in the intervention (Forms 2–4), including grades prior to the intervention, baseline and endline surveys, final grades, including national final exam (MSCE) scores in 2018, as well as 2019 and 2020, and a long-term follow-up survey (2023–2025). Forms 2–4 students are tracked through their schooling trajectories: taking the MSCE in their final secondary year, a gap year, and two years post-graduation when students may enroll in higher education. Since data collection for Form 2 students began in December 2023 (3 surveys), we observe all cohorts for up to three years after secondary school completion.

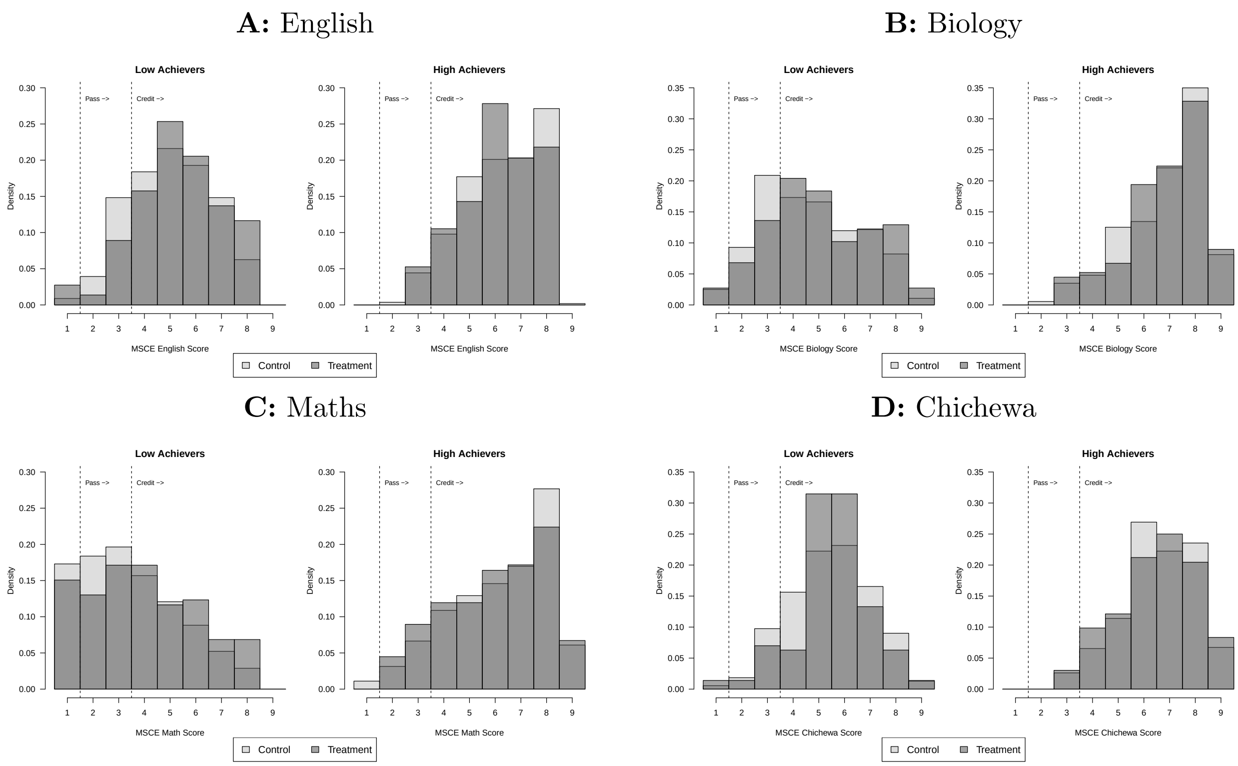
Appendix Figure A3: Comparison of Career Aspirations and Actual Programs of Study



A: Career Aspirations at Endline **B:** Careers based on Programs of Study

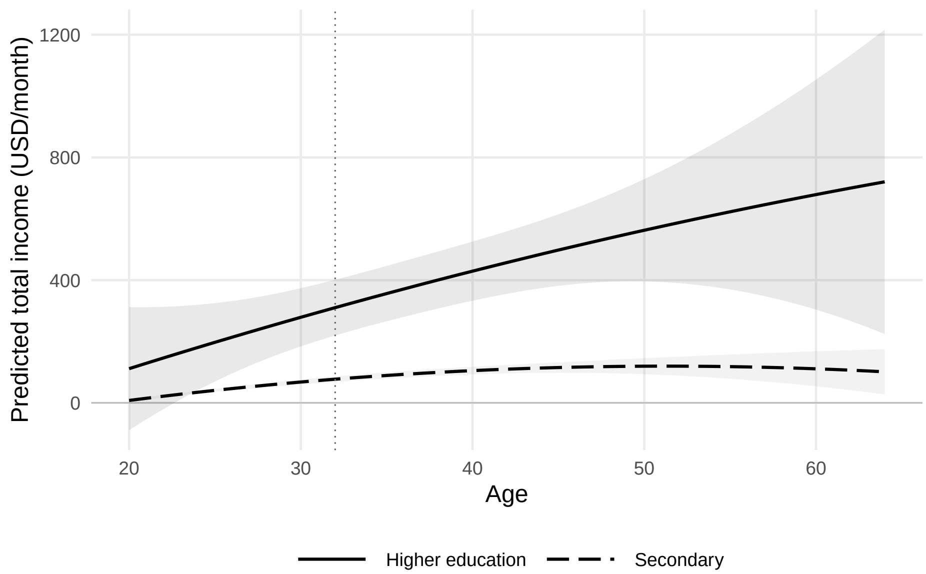
Notes: Panel A is from Derksen et al. (2022). It shows the *percentage* of students by career aspiration in the end-of-year survey. The sample is treatment students and the subsample of control students with supplementary surveys. Panel B shows the *number* of students who enrolled in a higher education program within three years after graduation, which leads to these different careers. Other Careers in Panel B include programs in Natural Science, Public Health, and Social Science.

Appendix Figure A4: Distributions of National Exam Scores by Subject



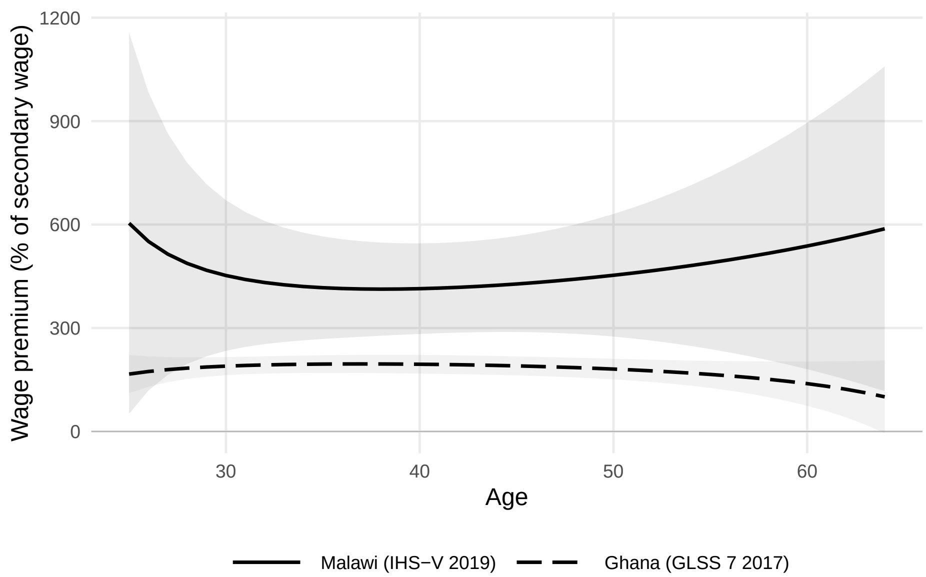
Notes: Panels A–D show histograms of raw MSCE scores by subject. In all panels, the left side shows low achievers and the right side shows high achievers, by treatment status. Scores were reversed so that higher values indicate better performance (10 minus raw score). Vertical lines indicate the minimum score required to pass and receive a credit. Low (high) achievers had average baseline exam scores in English and Biology below (above) the median.

Appendix Figure A5: Predicted Total Income by Age and Education



Notes: Predicted total monthly income by age (2018 USD, deflated from the 2019 survey values using the US Consumer Price Index), from separate quadratic-in-age regressions, with no additional controls, for adults whose highest completed qualification is secondary education (MSCE or A-level; $N = 2{,}305$) and adults who completed higher education ($N = 548$), estimated from the Fifth Integrated Household Survey (Malawi IHS5, 2019). The dependent variable is unconditional monthly income in USD (zero for non-earners), combining four individual-level sources: wage employment, casual day labor, self-employment income, and the imputed value of family farm labor at the median daily wage for casual labor. Shaded bands are 95% confidence intervals based on robust standard errors. The dotted vertical line marks age 32, the start of the projection window used in the cost-effectiveness benefit calculation.

Appendix Figure A6: Higher-Education Wage Premium by Age in Malawi vs. Ghana



Notes: Regression-adjusted higher-education wage premium as a percent of the predicted secondary-completer wage at each age, in Malawi (N = 2,853; Malawi IHS5, 2019) and Ghana (N = 6,215; Ghana GLSS 7, 2017). To keep the two countries comparable, the outcome is main-job wage earnings only (USD/month, zero for non-earners), excluding self-employment and farm income, which the GLSS reports only at the household level. In each country, the dollar premium is the coefficient on higher education from a regression interacting a higher-education indicator with age and age squared, controlling for gender, religion, father's and mother's education, birthplace fixed effects, and survey-month fixed effects; it is divided by the predicted secondary-completer wage at the same age from a quadratic-in-age fit on the secondary-only sample. Shaded bands are 95% confidence intervals based on robust standard errors, scaled by the same base.

B Spillover Analysis

B.1 Social Network Construction and Spillover Estimation

In both baseline and end-of-year surveys, students were asked to nominate peers from their school in response to a series of social network questions. Each question asked the respondent to list the identification numbers of peers with whom they shared a particular type of relationship. The 14 baseline questions and 12 end-of-year questions are listed below.

Question	Baseline	End-of-year
Best friend	✓	✓
Study partners	✓	✓
Borrowed money from	✓	✓
Borrowed things from	✓	✓
Given a gift to	✓	✓
Talk about personal topics	✓	✓
Attend religious activities with	✓	✓
Talk about entertainment	✓	✓
Siblings at school	✓	
Relatives at school	✓	
Ask about class topics	✓	✓
Ask about news	✓	✓
Ask about health	✓	✓
Ask about school activities	✓	✓
Total questions	14	12

We construct undirected networks separately for each wave as follows:

1. **Directed edge list.** For each respondent and each question, we extract the list of nominated peer identification numbers. Each nomination produces a directed edge from the respondent to the nominee. Self-nominations are dropped. This yields 60,668 directed edges across both survey waves (31,982 at the baseline survey, 28,686 in the end-of-year survey).

2. **Undirected network.** We collapse across all question types within each wave. An undirected edge exists between students i and j if either i nominated j or j nominated i on any question. Edge weights reflect the total number of nominations (across all questions and both directions).

3. **Sample restriction.** We restrict the network to students in the study sample (i.e., those appearing in the survey data with non-missing treatment status), dropping edges

where either endpoint is outside the sample.

Summary statistics for the resulting networks are reported below.

	Baseline	End-of-year
Network level		
Undirected edges	13,607	12,756
Unique nodes	1,508	1,495
Node-level degree		
Mean	18.05	17.06
Median	17	16
Min	2	1
Max	101	53
Number of treated links		
Mean	3.64	3.63
Median	3	3
Min	0	0
Max	20	16
Share of links that are treated		
Mean	0.200	0.212
Median	0.200	0.206

The correlation between baseline and end-of-year values of the number of treated links is 0.442.

To examine whether peer exposure to treated students affects exam performance, we estimate the following specification for each MSCE outcome y_i:

$$y_i = \beta_1 \text{Treatment}_i + \beta_2(\text{Treatment}_i \times \text{TreatedFriends}_i) + \beta_3(\text{Treatment}_i \times \text{Friends}_i) \\ + \beta_4 \text{TreatedFriends}_i + \beta_5 \text{Friends}_i + \boldsymbol{\chi}\mathbf{X}_i + \sigma_s + \varepsilon_i$$

where TreatedFriends_i is the number of treated peers in student i's baseline network, Friends_i is the student's total baseline degree (number of links), $\mathbf{X}_i$ includes baseline English score and an indicator for missing baseline score, and σ_s are randomization strata fixed effects. Standard errors are heteroskedasticity-robust.

Total baseline degree is positively and significantly associated with exam scores across most subjects (Table A15), suggesting that better-connected students perform better on average. The number of treated friends does not significantly impact outcomes, with the exception of mathematics scores, where we see a small negative but significant spillover effect on control students, which is attenuated and not significant for treated students (the p-value of the combination of Treatment × treated friends and Treated friends is 0.223). The main effect estimates are near-identical to the estimates in Panel A of Table 1.

B.2 Direct Effect with Endogenous Network Spillovers

We also estimate a model that allows for both spillovers and endogenous network change (Comola and Prina, 2021). To fit this model, we must make the following changes to the specification. First, the Comola and Prina (2021) model requires a first-differences specification with raw outcomes. We therefore define the outcome as a learning gain $\Delta y_i = y_{1i} - y_{0i}$, where y_{0i} is the student's raw baseline school-exam grade (0–100) and y_{1i} is the MSCE grade rescaled onto a 0–100 "percent of max" axis. Both sides of the difference are therefore in percentage-point units, and Δy should be interpreted as a percentage point gain on this scale. Second, we must define networks within clusters. We therefore use the same list of link types as above, restricted to within-classroom dyads (36 clusters).

Following Comola and Prina (2021) equation (6), we estimate the dynamic peer-effects specification

$$\Delta y_i \;=\; \delta_1 (G_0 \Delta y)_i + \delta_2 (\Delta G\, y_1)_i + \beta\, \mathrm{ITT}_i + \gamma_1 (G_0 \cdot \mathrm{ITT})_i + \gamma_2 (\Delta G \cdot \mathrm{ITT})_i + \sigma_s + \varepsilon_i,$$

where G_0 and G_1 are the (semi row-standardized) within-class undirected interaction matrices at baseline and endline, $\Delta G = G_1 - G_0$, and σ_s are randomization-strata fixed effects. We estimate this by 2SLS, with two endogenous regressors: the outcome peer effect $G_0 \Delta y$, and the combined network-change term $\Delta G\, y_1 = \Delta G\, y_0 + \Delta G\, \Delta y$. Instruments are the second- and third-order products of G_0 and ΔG applied to ITT (twelve instruments in total). The reported ITT coefficient $\widetilde{J}_{itt} \equiv \beta$ is the direct treatment effect net of peer-effect channels. Standard errors are cluster-robust at the school×form×class level (36 clusters).

For completeness, we also report a no-peer-effects benchmark (OLS of Δy on ITT and strata fixed effects), corresponding to column (1) of Comola and Prina (2021) Table 1, and the same benchmark run separately on low-achiever and high-achiever subsamples.

Table A16 reports the results. The coefficients must be interpreted differently from other tables in the paper as they reflect the impact on a raw score (between 0 and 100) first-difference. This first-difference specification has lower statistical power than the ANCOVA specification used throughout the paper. However, the estimated effects are qualitatively consistent with the main findings, with positive average treatment effects and significant impacts for low achievers only. Importantly, the estimates of the direct treatment effect are very similar with and without taking endogenous peer effects into account. For example, the average treatment effect for the aggregate score is 1.31 points as estimated by OLS, and 1.21 points under Comola and Prina (2021), both significant at the 10 percent level.

B.3 National MSCE Analysis

We use administrative records from the Malawi School Certificate of Education (MSCE) examination for all secondary schools in Malawi from 2015 to 2024. These records include individual-level scores in all subjects taken, which we use to construct measures of exam performance.

Data from the four study schools is supplemented with data from approximately 1,470 other secondary schools that appear in the national records.

As in Section 3, we construct the following outcome variables from the administrative records:

- **Subject scores** (columns 1–5 of Table A17). Raw MSCE scores in English, Biology, Mathematics, Chichewa, and an aggregate score (sum of best 6 subjects). Scores are reversed so that higher values indicate better performance, then standardized (z-scored) relative to the non-study school distribution within each exam year. Note that this normalization is different from the normalization used for the main results, where we normalize with respect to control students within the study schools.
- **Pass** (column 6) and **Qualified for university** (column 7). Defined as in Section 2.1, but conditional on taking at least one exam (since the administrative records only include students who sat the MSCE). In the main analysis, these indicators are defined unconditionally, with students who did not sit the exam coded as zero.
- **Number of exam takers** (column 8). Total number of students sitting the MSCE at each school in each year. This regression is estimated at the school-year level rather than the individual level.

We estimate a two-way fixed effect specification comparing the four study schools to all other schools, before, during, and after the intervention period. For individual-level outcomes (columns 1–7):

$$y_{ist} = \beta_1 \,\mathrm{TreatedSchool}_s \times \mathrm{InterventionYears}_t + \beta_2 \,\mathrm{TreatedStudent}_i + \gamma \,\mathrm{Male}_i + \alpha_s + \delta_{t\times b} + \varepsilon_{ist}$$

where $\mathrm{TreatedSchool}_s \times \mathrm{InterventionYears}_t$ equals one for all students at study schools taking their MSCE during the intervention years (2018–2020), $\mathrm{TreatedStudent}_i$ equals one for individually randomized treated students within study schools, α_s are school fixed effects, and

$\delta_{t\times b}$ are exam year $\times$ boarding school fixed effects (allowing for different trends in boarding and non-boarding schools). Standard errors are clustered at the school level.

The coefficient β_1 captures the school-level treatment effect—the average difference between study schools and comparison schools during the intervention period, net of school and time fixed effects. This reflects both the direct effect of being in a study school and any school-wide spillovers from the intervention. The coefficient β_2 captures the additional effect of individual treatment assignment within study schools. For the number of exam takers (Column 8), we estimate the same specification at the school-year level without the individual treatment variable.

A key challenge for inference is that treatment is assigned at two levels: school-level (4 treated schools) and individual-level (approximately 20% of students within study schools). With so few treated clusters, cluster-robust standard errors and Wald F-tests have poor finite-sample properties.

To address this, we use permutation inference (analogous to randomization inference) that mirrors the actual two-level study design (Young, 2019). The procedure is as follows:

1. Estimate the actual test statistics from the data (the t-statistics for β_1 and β_2, and the F-statistics for joint significance of pre- or post-trend coefficients).

2. Randomly reassign "treated school" status to 4 schools drawn from the pool of non-study schools. Within those schools' intervention-period students, randomly assign individually treated students stratified by year, matching the actual within-cohort treatment fraction (approximately 20%). Note that we do not have baseline data required to match the study stratification procedure exactly. Re-estimate the model and compute the same test statistics under this placebo assignment.

3. Repeat step 2 for 10,000 permutations to build the distribution of the test statistics under the null hypothesis of no treatment effect.

4. The permutation p-value is the fraction of permuted test statistics that exceed the actual test statistic, comparing t-statistics in absolute value (a two-sided test); F-statistics are non-negative and compared directly.

Significance stars for both the school-level treatment effect (β_1) and the individual treatment effect (β_2), as well as p-values for the pre- and post-trend tests, are based on this permutation procedure.

To assess the validity of the parallel trends assumption, we augment the main specification with year-specific indicators for the study schools in the pre-treatment period (2015–2016, with 2017 as the reference year) and the post-treatment period (2021–2024). We then test the joint significance of the pre-treatment indicators (pre-trends test) and the post-treatment indicators (post-trends test) using the permutation procedure described above.

The individually randomized treatment (β_2) has positive point estimates for the aggregate score (0.116 SD, $p = 0.26$), and the probability of qualifying for university admission (5.9 percentage points, $p = 0.06$). The school-level treatment effect (β_1) is not statistically significant for any outcome, and the impacts on aggregate score (-0.019 SD, $p = 0.26$) and university admission (1.0 percentage points, $p = 0.83$) are close to zero. All pre-trends permutation p-values exceed 0.6, providing no evidence of differential pre-trends between study schools and comparison schools. All post-trends permutation p-values also exceed 0.6, indicating that the study schools do not diverge from comparison schools after the intervention period ends. The number of exam takers at study schools during the intervention period is similar to comparison schools (-2.1, $p = 0.31$), suggesting that the intervention did not affect exam-taking rates.

C Cost-Effectiveness Calculations

This appendix provides detailed calculations supporting the cost-effectiveness claims in the main text. We first present the calculations for our intervention. Then, we calculate the marginal value of public funds (MVPF) and benefit-cost ratio (BCR) following the framework of Hendren and Sprung-Keyser (2020), and compare cost-effectiveness to two benchmark studies. Finally, we assess robustness to alternative assumptions about earnings returns and wage growth. We normalize all costs and benefits to 2018 USD and apply a 5% annual discount rate.

C.1 Intervention Cost and Impact

The intervention cost approximately \$36 per student for one school year, including project management, digital library staff, devices, and data packages (Derksen et al., 2022). The key treatment effects used in the calculations below are:[17]

- Aggregate Malawi School Certificate of Education (MSCE) score: 0.14 standard deviations
- Probability of enrolling in higher education: 6.5 percentage points
- Years of higher education (within three years of graduation): 0.08 (Table A20)

The intervention improves aggregate MSCE scores by 0.14 SD at a cost of \$36 per student:

$$\frac{\$36}{0.14/0.1} = \frac{\$36}{1.4} \approx \$26 \text{ per } 0.1 \text{ SD}$$

The treatment effect on higher education enrollment is 6.5 percentage points per \$36 spent. The cost per additional higher education enrollee is therefore:

$$\frac{\$36}{0.065} = \$554 \text{ per additional enrollee}$$

The treatment effect on years of higher education is 0.08 years per \$36:

$$\frac{0.08}{\$36} \times \$100 = 0.22 \text{ years per } \$100$$

This is likely a lower bound, as we observe only the first three years after secondary school graduation. Some treated students are still enrolled or will enroll in additional years

[17] Throughout this appendix, displayed formulas use rounded intermediate values for exposition; reported results use precise inputs throughout.

of higher education beyond our observation window. Among students in our sample who enrolled in higher education and whom we observe at least one year after enrollment, 91 percent are still enrolled one year later. Assuming a nine percent annual dropout rate over a three-year degree, each enrollee attends $1 + 0.91 + 0.91^2 = 2.74$ years in expectation, and $0.91^2 = 83$ percent complete the degree. With a 6.5 percentage point increase in enrollment, this implies a predicted increase in average years of higher education of $0.065 \times 2.74 = 0.18$ years. This is an underestimate if additional students enroll or re-enroll beyond the 3-year window, and an overestimate if the dropout rate increases over time after the first year.

C.2 MVPF and BCR

Costs

The intervention generates three types of costs:

1. **Intervention costs (C_I = \$36):** The cost of the intervention per student for one school year, including project management, digital library staff, devices, and data packages (Derksen et al., 2022).

2. **Direct educational costs (C_{ED}):** The intervention increases higher education enrollment, generating additional educational expenses borne by households (C_{ED}^{pri}) and the government (C_{ED}^{govt}).

   Private cost. As of 2018, tuition fees at the University of Malawi were approximately \$480 per year.

   Public cost. We estimate government expenditure per tertiary student by combining three sources. Malawi's 2018 GDP was \$9.88 billion, of which 3.32 percent was spent on education (World Bank, 2024). Approximately 25 percent of the education sector budget was allocated to tertiary education in 2021/22, prior to the creation of the Ministry of Higher Education (UNICEF, 2025). Gross tertiary enrollment is approximately three percent (UNESCO, 2025), implying approximately 90,000 students. This yields a public expenditure of approximately \$911 per student per year.[18] This is likely an overestimate of the marginal cost due to large fixed costs.

   Educational costs are incurred for every year attended, including by students who do not complete the degree. We therefore weight annual costs by the predicted increase in average years of higher education (0.18 years): $C_{ED}^{pri} = 0.18 \times \$480 = \$86$ and $C_{ED}^{govt} = 0.18 \times \$911 = \$163$.

[18] Calculated as $0.25 \times 0.0332 \times \$9.88$ billion$/90{,}000 \approx \$911$.

3. **Opportunity costs of education (C_{EO} = \$12):** Based on our survey data, the difference in earnings within three years of graduation between those who never enrolled in higher education and those who completed at least one year implies forgone earnings of approximately \$66 per year, converting nominal kwacha amounts at each year's average official exchange rate and deflating to 2018 USD using the US CPI. This figure is higher than average employment earnings for secondary school graduates aged 18 to 21 in the nationally representative Fifth Integrated Household Survey (Malawi IHS5, 2019) data (\$17 per year). The opportunity cost attributable to the intervention is $C_{EO} = 0.18 \times \$66 = \$12$.

Labor market benefits

We estimate labor market returns by applying causal estimates from the Ghanaian scholarship study of Duflo et al. (2025) via the surrogate approach of Athey et al. (2019) (as in Ashraf et al. 2025). We use the direct reduced-form causal effect of the Ghanaian scholarship on *total* earnings. By 2023, the program raised total earnings (including wages, casual labor and self-employment) by 12.5 percent (Duflo et al., 2025). This gain is likely attributable to higher education: there is no earnings effect for men, who gain no additional higher education, and two-thirds of the women's gain comes from access to government jobs (teaching and nursing) that require higher education.

We therefore scale the estimated causal effect from Duflo et al. (2025) by the ratio of our higher-education enrollment effect to theirs, $(0.065/0.077) \times 12.5\% = 10.6\%$. We assume participants earn this 10.6 percent premium at each age over a working life from age 32 to 60. Starting the stream at age 32 matches the Ghanaian evidence: participants in Duflo et al. (2025) averaged 17 years of age at their 2008 baseline, so the earnings effect measured in their 2023 wave is an effect at approximately age 32, while their 2019 wave (approximately age 28) shows no earnings effect yet.

We apply the earnings effect to the age-income profile of secondary-educated Malawians, estimated from the nationally representative Fifth Integrated Household Survey (Malawi IHS5, 2019). Among adults whose highest completed qualification is secondary education, we regress total monthly income on a quadratic in age, fitting the regression on the full working-age sample (ages 15 to 64) and taking predicted income at each age over the projection window of 32 to 60. The income measure combines four individual-level sources: public and private-sector wage employment, self-employment income, casual day labor, and the imputed value of family farm labor, based on the median daily wage for casual labor. The measure includes zeros for those without earnings, so it captures both the extensive (employment access) and intensive (earnings conditional on work) margins. Our measure is directly comparable to Duflo et al. (2025), who likewise aggregate earnings across all activities, including casual labor and self-employment, with zeros for non-earners. Incomes are observed in 2019 (the year of the IHS5 fieldwork) and deflated to 2018 USD using the US Consumer Price Index. Predicted income for those with secondary education rises from \$77 per month at age 32 to \$119 at age 50. We plot predicted income for both secondary and higher-educated Malawians in Figure A5; the higher-education curve is for descriptive purposes only and is not used in the calculations.

Earnings are indexed in real terms to 2019 levels, grown forward at Malawi's average annual real GDP growth rate over the pre-pandemic decade (2010–2019) of 4.4% per year (World Bank, 2024), and discounted to 2018 at 5%. For simplicity we anchor the projection on a representative participant who graduates from secondary school at age 18 in 2018, the graduation year of the oldest cohort, so that age a is reached in calendar year $t(a) = 2018 + (a - 18)$. The gross earnings gain per participant—the total benefit—is

$$B = 0.106 \times 12 \sum_{a=32}^{60} \hat{y}_{sec}(a) \frac{(1.044)^{t(a)-2019}}{(1.05)^{t(a)-2018}} = \$3{,}237,$$

where $\hat{y}_{sec}(a)$ is the predicted monthly income of a secondary completer at age a, in 2018 USD (Figure A5). Anchoring the younger cohorts to their actual graduation years (2019 and 2020) would delay the earnings stream and reduce the discounted benefit by less than one percent per year of delay, as wage growth (4.4 percent) nearly offsets the discount rate (5 percent).

We apply Malawi's 2018 Pay As You Earn (PAYE) tax schedule to public-sector earnings and assume no tax on private-sector earnings, as most private-sector employment is informal (Malawi Revenue Authority, 2018); at the mean government wage, the effective tax rate is $\tau = 22.4$ percent.[19] Since about two-thirds of the earnings gain comes from public-sector employment (Duflo et al., 2025), the taxable public-sector portion of the gain is $B^{pub} = \frac{2}{3}B$. This generates tax revenue $R = \tau \times B^{pub} = \$483$, and participants keep the net-of-tax benefit $B_N = B - R = \$2{,}754$.

[19] Malawi Revenue Authority PAYE brackets (effective July 2018): first K35,000 at 0%, next K5,000 at 15%, excess at 30%. At the mean government wage of K147,750 per month, tax = (0+750+32,325)/147,750 = 22.4%.

Results

The MVPF is defined as the ratio of participants' willingness to pay (WTP) for the intervention to the net cost borne by the government (Hendren and Sprung-Keyser, 2020):

$$MVPF = \frac{B_N - C_{ED}^{pri} - C_{EO}}{C_I + C_{ED}^{govt} - R}$$

The benefit-cost ratio is:

$$BCR = \frac{B}{C_I + C_{ED}^{pri} + C_{ED}^{govt} + C_{EO}}$$

The total benefit is $B = \$3{,}237$, against total costs of $\$36 + \$86 + \$163 + \$12 = \$296$, giving a BCR of 10.9. Because the \$483 in tax revenue exceeds the government's direct costs of \$199 ($C_I + C_{ED}^{govt}$), the net public cost is negative: the MVPF is infinite, and the intervention more than pays for itself in fiscal terms.

The cost-effectiveness of the intervention is comparable to that of two secondary school interventions with the highest documented returns: the Ghanaian secondary school scholarship program studied by Duflo et al. (2025), which randomly assigned full scholarships to youth who had gained admission to senior high school but could not afford to enroll, and the Zambian negotiation-skills program (Ashraf et al., 2020, 2025). Table A18 summarizes the comparison.

Scenario analysis

The estimates from Duflo et al. (2025) may not map perfectly to our setting for two reasons. First, marginal higher education enrollees may be different under different interventions, and the effects on earnings and employment sector may also be different. Second, the higher-education wage premium appears to be approximately twice as high in Malawi as in Ghana (Figure A6, based on data from Malawi IHS5 2019 and Ghana GLSS 7 2017), so the Ghana-scaled effect may understate the gain.

Table A19 reports the BCR and MVPF under three premium scenarios: the Ghana-scaled effect (10.6 percent), half that effect (5.3 percent), and twice that effect (21.1 percent). Because the earnings premium accrues decades into the future, results are also sensitive to assumed real wage growth; we cross each premium scenario with our main assumption of 4.4 percent growth (the 2010–2019 average) and a lower 2.9 percent (the 2015–2024 average, incorporating the slower post-pandemic years).

Across scenarios, the BCR ranges from 3.7 to 21.8, and the MVPF from 24 to infinite.

All scenarios remain conservative in one respect: we do not include intergenerational benefits to participants' children documented by Duflo et al. (2026), as these may be partially attributed to increases in secondary (as opposed to higher) education.

D Example of an Examination in Mathematics

NAME ____

2019 MSCE EXAMINATIONS
MATHEMATICS
Subject Number: M131/II
Time Allowed: $2\frac{1}{2}$ hours
PAPER II (100 marks)

Instructions

- Answer all six questions in Section A and any four questions in Section B.
- The maximum number of marks for each answer is indicated against each question.
- Write your answers in the spaces provided on the question paper.
- Calculators may be used.
- All working must be clearly shown.
- Write your NAME at the top of each page of your question paper.

SECTION A (60 marks) — Answer all six questions

Question 1

a. A chord subtends an angle of 110° at the centre of a circle of radius 10 cm. Calculate the length of the chord. (5 marks)

b. Convert 126_7 to base 10. (5 marks)

Question 2

a. Using a ruler, a pencil and compasses only, construct in the same diagram:

- a triangle ABC in which $AB = 7$ cm, $BC = 13$ cm and angle $ABC = 60$.
- the inscribed circle of this triangle.

(7 marks)

b. Find the gradient of a line joining the coordinates $(-10, 3)$ and $(-2, -1)$. (5 marks)

Question 3

a. Factorise $15 - 13x + 2x^2$. (5 marks)

b. If $f(x) = \frac{x+1}{2x-3}$, find $f(-1)$. (4 marks)

Question 4

a. In a building there are 24 cylindrical pillars. The radius of each pillar is 28 cm and height is 4 m. Find the total cost of painting the curved surface area of all pillars at the rate of K8,000 per m^2. (6 marks)

b. Calculate the value of angle b. (5 marks)

Question 5

a. Calculate the number of terms in the geometric progression: $2, 4, 8, \ldots, 128$. (5 marks)

Question 6 Five schools sent some students to a conference. One of the schools sent both boys and girls; this school sent 16 boys. The ratio of the number of boys it sent to the number of girls it sent was 1 : 2. The other 4 schools sent only girls. Each of the 5 schools sent the same number of students. Work out the total number of students sent to the conference by these 5 schools. (5 marks)

Question 7

a. Given that X and Y are sets containing birds where

$$X = \{\text{njiwa, pumbwa, kakowa, nkhwangwala}\}$$

$$Y = \{\text{nkhunda, timba, kakowa, njiwa, nkhwali, nkhanga}\}$$

- Find $X \cup Y$.
- Find $n(X \cap Y)$.

(5 marks)

b. Simplify $3^4x^5 \div 3x^{-4}$. (4 marks)

SECTION B (40 marks) — Answer any four questions

Question 8 The probability of Monday being dry is 0.6. If Monday is dry, the probability of Tuesday being dry is 0.8. If Monday is wet, the probability of Tuesday being dry is 0.4.

a. Show this in a tree diagram.

b. What is the probability of both days being dry?

c. What is the probability of both days being wet?

d. What is the probability of exactly one dry day?

(10 marks)

Question 9

a. Triangle ABC has vertices $A(1,2)$, $B(1,5)$ and $C(3,5)$. It is rotated clockwise through 90° about $(2,0)$. Draw the triangle ABC and its image after the rotation.

b. Solve the simultaneous equations:

$$xy = 4$$

$$x + y = 5$$

(10 marks)

Question 10

a. Find the equation of the straight line PQ.

b. Complete the table of values for the equation $y = 27 + 3x - 2x^2$.

c. Draw the graph of $y = 27 + 3x - 2x^2$ for values of x from -3 to 4.

d. Write down the x-coordinates of the points where the line PQ intersects $y = 27 + 3x - 2x^2$.

(10 marks)

Question 11 A rectangular water tank has dimensions: $AB = 10$ cm, $BC = 7$ cm, and $BF = 5$ cm.

i. Calculate the length of DF leaving your answer in its simplest surd form.

ii. Calculate the angle which DF makes with the plane ABCD.

(10 marks)

Question 12 A town council plans to build a car park for x cars and y buses. Cars are allowed 10 m^2 of ground space and buses 20 m^2. There is only 500 m^2 available. The park can hold a maximum of 40 vehicles at a time, and only a maximum of 15 buses at a time.

a. Write down three inequalities other than $x \geq 0$ and $y \geq 0$.

b. On graph paper, show the region that satisfies the given conditions.

c. If the parking charges are K50 for cars and K200 for buses per day, find the number of vehicles of each type that should be parked to obtain maximum income.

d. Calculate the maximum income.

(10 marks)

Question 13 A flag post stands on top of a church tower. From a point on level ground, the angles of elevation of the top and bottom of the flag post are 40° and 30° respectively. The flag post is 6 m long.

a. Calculate the height of the church tower.

b. Calculate the distance of the observer from the church tower.

c. Use sine rule to calculate the distance from the observation point to the flag post.

(10 marks)

End of Question Paper