Populated Pre-Analysis Plan for Guatemalan Unified Assistance and Resources for Development, Inspiration and Academic Navigation (GUARDIAN)*

Carla Coccia† Eric Edmonds‡ Martina Jakob§ Mauricio Romero¶

January 27, 2026

Abstract

This PAP is for AEA registry 13889 (https://www.socialscienceregistry.org/trials/13889). Evidence from developed countries shows that community schools, mentoring, and guidance counseling programs effectively reduce school dropout. This success is attributed to their ability to provide crucial social support and valuable information. Yet, evidence for developing countries remains limited and the key mechanisms driving the success of these programs remain to be disentangled. To answer these questions, the GUARDIAN project will examine two types of interventions in Quiché, Guatemala, through a randomized controlled trial (RCT). The two interventions are a school-based mentoring and information program and an information-only campaign. The trial also includes a control group. By comparing the outcomes of the combined program and the information-only campaign, the project aims to identify the relative importance of social support versus informational content in reducing school dropout rates in a low-income setting. This version of the PAP is revised to reflect the actual data collected at endline, which differs from our previous PAP because of a change in funding.

*The project, funded by the United States Department of Labor under cooperative agreement number 4K75IL039379-01-02, is implemented by Innovations for Poverty Action (IPA) in collaboration with local stakeholders including the Ministry of Education of Guatemala and the NGO 'Proyecto de Desarrollo Santiago' (PRODESSA). Additional support for PRODESSA on this project is provided by a grant from the U.S. Department of State's Human Trafficking Research Initiative. 100 percent of the total costs of the project or program is financed with USG federal funds, for a total of $1.7 million dollars. This material does not necessarily reflect the views or policies of the United States Department of Labor or the United States Department of State, nor does mention of trade names, commercial products, or organizations imply endorsement by the United States Government. GUARDIAN received IRB approval from Human Subjects Committee for Innovations for Poverty Action IRB-USA #16288. Rosa Miranda Santa Cruz, Dalma Villanueva and Daniel Hernandez provided excellent field management and research assistance through Innovations for Poverty Action (IPA). Romero gratefully acknowledges financial support from the Asociación Mexicana de Cultura, A.C. Edmonds is grateful to Ariana Rodriguez Bruzon for exceptional research assistance.

†University of Bern; E-mail: carla.coccia@unibe.ch

‡Dartmouth College; E-mail: eric.v.edmonds@dartmouth.edu

§University of Zurich; E-mail: martina.jakob@econ.uzh.ch

¶ITAM; J-PAL; E-mail: mtromero@itam.mx

1 Introduction

1.1 Motivation

GUARDIAN was being implemented in 15 municipalities in Quiché. Quiché has one of the highest migration rates in Guatemala, with 11% of the country's international migrants originating from this region (International Organization for Migration (IOM) and United Nations Population Fund (UNFPA), 2021). Quiché exhibits one of the lowest school transition rates nationwide, with only 45% of students staying in school after grade 6.[1]

This study aims to understand the role of student beliefs and capacity to act on those beliefs in school dropout, child employment, and child migration decisions. We randomize two interventions aimed at students at the end of primary school, grade 6. Randomization is at the school level. An information treatment provides information aimed at updating beliefs about the schooling, work, and migration environment, including legal constraints and expected consequences. A mentoring treatment provides the same information to update beliefs and additionally attempts to relax capacity constraints that determine whether those beliefs can be acted upon.

1.2 Research Questions

The research team aims to address the following pivotal questions:

- What are the effects of the mentoring and information programs on the transition from primary to secondary level in sixth-grade students in Quiché?
- What are the effects of the programs on the allocation of time and child employment decisions of sixth-grade students and their families?
- What are the effects of the programs on the migration decision of sixth-grade students and their families?

1.3 Theory of Change

Children whose parents have little or no formal education, living in communities with low educational attainment, often face constraints that limit their ability to transition successfully beyond primary school. In such settings, children may hold incomplete or inaccurate beliefs about the returns to education, the costs of continued schooling, and the pathways required to remain enrolled. These belief gaps are compounded in contexts where individuals who benefit from higher education tend to migrate,

[1] Authors' calculations from Ministry of Education data in Haimovich et al. (2021).

making it difficult for children to observe local returns to schooling or to understand how education translates into future opportunities. This study is motivated by the hypothesis that both beliefs about education and the capacity to act on those beliefs play central roles in schooling, child labor, and migration decisions.

A substantial body of evidence demonstrates that informational interventions can shape beliefs relevant for educational investment. Jensen (2010) shows that providing middle school boys in low-income areas of the Dominican Republic with information on the returns to education significantly increased school participation, attributing initial underinvestment to limited exposure to information in communities where educated individuals move away. Similarly, Avitabile and De Hoyos (2018) report positive impacts of educational information in Mexico, while cautioning that financial and other constraints may limit families' ability to act on updated beliefs. In Chile, Dinkelman and Martínez A (2014) find that correcting misperceptions about the costs of education increases school attendance. In the United States, Bettinger et al. (2012) emphasize the importance of informing students from low-participation backgrounds about how to continue their education. Building on this literature, our informational intervention aims to update students' and caregivers' beliefs about the returns to education, the costs of continued schooling, and the concrete pathways linking education to future opportunities, including education-to-career pathways shown to motivate schooling in higher-income contexts (McGuigan et al., 2016).

The informational intervention also targets beliefs related to child labor and human trafficking. Students and their families receive information on Guatemalan laws governing child labor and on the dangers associated with human trafficking. Research by Margaret Boittin (2016) finds that raising awareness about trafficking risks was effective in Nepal, with potential victims absorbing the information provided. Although many anti-child labor programs incorporate informational campaigns, evidence on their direct effects on children's time use remains limited.

We think of the information intervention then as both updating beliefs about the desirability and permissibility of different decisions made at the end of primary. By updating beliefs about the desirability and permissibility of different activities, the informational intervention seeks to influence how families evaluate schooling, work, and migration options.

The central question therefore becomes whether children and families can act on the beliefs they hold. To address this, the study includes a mentoring component that builds on the informational intervention by both reinforcing belief updates and

relaxing constraints that limit the capacity to act on those beliefs. The mentoring arm delivers the same informational content as the information-only intervention, but for students, it does so in a more interactive manner, potentially strengthening belief formation through reinforcement and salience. In addition, mentoring aims to expand students' capacity to translate beliefs into sustained action by supporting effort, planning, accountability, and access to social and institutional support.

The mentoring intervention includes role-playing exercises and life skills development designed to help students navigate schooling-related decisions and challenges, strengthen their ability to follow through on plans, and engage effectively with adults and institutions. Similar life skills programs in Tanzania, Bangladesh, and Ethiopia have improved the mental health of adolescent girls (Shah et al., 2024). Programs designed to help students advocate for themselves have also increased school attendance in Zambia (Ashraf et al., 2020) and India (Edmonds et al., 2023). In this sense, mentoring is conceptualized not as altering preferences, but as increasing the feasibility of acting on existing beliefs by reducing constraints related to effort provision, implementation, coordination, and persistence.

Our mentoring program integrates life skills with practical information, mirroring elements of in-school guidance and counseling programs common in higher-income countries. Evidence suggests that such guidance programs have contributed to the expansion of education in those contexts (Carrell and Hoekstra, 2014). Through this combined approach, the study examines whether reinforcing beliefs and relaxing capacity constraints together are more effective at shaping schooling, child labor, and migration outcomes than updating beliefs alone.

Figure 1 summarizes how we think these interventions will interact and influence school enrollment, child employment, and migration. We anticipate that the information treatment will change beliefs about the value of schooling and its alternatives if subjects have inaccurate beliefs and that altering those beliefs will impact schooling, employment, and migration.[2] The mentoring intervention contains the information treatment and may better reinforce the delivery of that information as well as change the capacity to act on changes in beliefs.

1.3.1 Consistency with the Registered Pre-Analysis Plan.

The only substantive change in how we are thinking about this theory of change is to put the beliefs-constraints packaging around the prior content and a few minor rewordings.

[2]Marriage is often an important consideration in early school termination in settings where dropout is at a later age. It would be extremely unusual in our setting to observe our 13 year olds marrying.

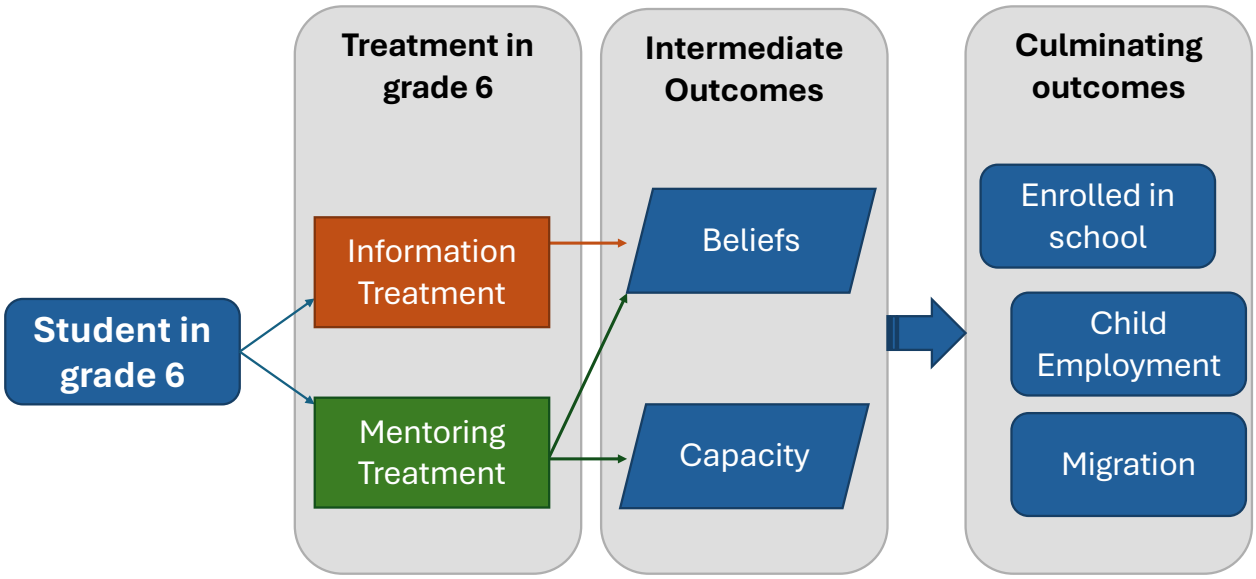


Figure 1: Theory of Change

This reframing is intended to clarify, not alter, the theory of change articulated in the registered Pre-Analysis Plan. The PAP motivates the informational intervention as addressing gaps in knowledge about the returns to education, schooling pathways, and risks associated with child labor and migration, and motivates the mentoring intervention as building on this information by supporting students in acting on what they learn. The beliefs versus capacity framework formalizes this distinction using standard economic language: informational content affects beliefs relevant for evaluating schooling, work, and migration choices, while mentoring additionally relaxes constraints that limit the ability to act on those beliefs. The mentoring arm continues to be conceptualized as delivering the same informational content as the information-only arm, augmented by additional reinforcement of the messaging, life skills, role-playing, and sustained support.

2 Research Strategy

2.1 Treatment

This study evaluates two closely related treatments.

The first treatment is similar to a guidance counselor program that would be commonplace in many schools in high-income places and shares many characteristics with so-called "community schools" programs that aim to support students holistically. A mentor is provided to each school. The mentor provides students with weekly in-class meetings for four months. Appendix 1 of this document summarizes the sessions contents. This program aims to empower children by providing knowledge and support focused on

personal and educational development. As a part of this program, the mentor provides students with information on the value of education, how to enroll in lower secondary school, and how to recognize child labor and human trafficking. We refer to this first treatment arm as "Mentoring and Information".

The second treatment arm provides the information of the first arm but without the empowerment and mentoring support of the first arm. We refer to this second arm as "Information". It includes 5 specific sessions: one for parents, one for teachers, a joint session for parents and teachers, and two for the students. It provides students, caregivers, and teachers with all the same take home material as the mentoring and information arm, as shown in Appendix 2. The key difference is that the mentors do not accompany students; the advantage is that it is much cheaper to implement.

2.2 Sampling

2.2.1 Sampling Frame

This study is conducted in the Guatemalan department of Quiché. From the list of all government schools in the department with grade 6 students, we eliminated schools from the sampling frame for the following reasons:

- They were in one of six municipalities where USAID was operating to avoid overburdening the schools with different programs and to estimate treatment effects against the prevalent "business-as-usual" in the country.
- In Department of Education enrollment records from 2013-2021, their average dropout rate at grade 6 was below 35 percent. We wanted to target the program to schools with high dropout rates.
- They were not accessible owing to security or transport infrastructure problems.

We then drew 225 schools for the study with a buffer of 2km around each selected school. ~~One of these schools refused to engage with the project and was dropped.~~[3]. Later, an additional 6 schools were added to the study as resources became available. Assuming approximately 10 participants per school, we anticipated a total of 2,310 students enrolled in the evaluation.

In Table 1, we compare schools in Quiché and our study sample to those in the rest of Guatemala. These data are from the school records held by the Ministry of

[3]This was imprecise and creating confusion. 1 school stopped participating in the study after the study began, but we continue to include them in our analysis as a non-complier.

Education. In Column 1, we describe schools in the whole of Guatemala which tend to have larger enrollments and more teachers than schools in Quiché. Compared to other schools in Quiché (Column 2), our study schools (Column 3) are more rural and bilingual with fewer teachers than other schools in Quiché or the rest of Guatemala. In Column 4, we report results of the hypothesis test that our study sample does not differ from the rest of Quiché. Our sampling rules above should result in different characteristics for our subsample, as we observe.

We follow the same structure in Table 2. We use the same Ministry of Education data as the previous table but look at individual student characteristics. These are not the students in our study as they are student characteristics associated with the schools well before our study. For the 2018 academic year, we have the most detailed data, including a government model designed to predict how at risk the students are for dropping out. Our study population is more male, older, more likely to have repeated a grade, has a lower GPA, and is considerably more at risk of dropping out (with a higher dropout rate) compared to students at other schools in Quiché and in the other parts of Guatemala. In these data, students in our sample schools were more than twice as likely to dropout at the end of grade 6 compared to students in other schools in Guatemala.

During the design of the midline survey, we learned that the baseline data collection team went to the wrong school for two schools (Aldea Xolcuay and Cantón Chitucur III). Random assignment and treatment occurred in the correct school rather than the baseline school. Hence, both schools will be included in the final analysis, but we are missing baseline data for those two schools.

2.2.2 Statistical Power

Power calculations were conducted around detecting an impact of the interventions on dropout. At the design stage, we assumed a significance level of 0.05, power of 0.8, a standard deviation of dropout of 0.5 (which implies a dropout rate of 0.5), an intra-cluster correlation of 0.3178 (Haimovich et al., 2021), and that control variables (gender, age, and strata fixed effects, which control for historical dropout rates and school size at the school-level) would explain 20 percent of the outcome's variance. With an initially planned sample size of 150 clusters (75 schools per treatment arm and 75 schools in the control) and 10 students per cluster, we estimated to be able to detect a 0.25 standard deviation effect on dropout. This corresponds to 12.5 percentage points or a 25 percent reduction in dropout. These were the original power calculations we conducted before we randomized, when we designed the experiment and before we had access to the baseline and to administrative data.

We were able to increase the sample size to 231 schools. We average 10 students per school. With this sample size and the other assumptions the same, our minimum detectable effect with no control variables is 11 percentage points or 0.23 standard deviations. If we use the actual dropout rate for our study schools (see Table 2), that is 65% with a standard deviation of 0.48 and the actual intra-cluster correlation (in 2021) for dropout of 0.14, we have a minimum detectable effect of 9 percentage points. In the available administrative records we have for our study schools, our regression controls account for 10 percent of the dropout variation, and the ICC for the dropout residuals is 0.07. These adjustments imply a minimum detectable effect of 8 percentage points (0.17σ). .

2.2.3 Assignment to Treatment

The 231 selected schools were grouped into 16 strata. 15 strata were formed based on the dropout rate and number of students enrolled. Schools were blocked into 5 bins for dropout and 3 bins for number of students, forming 15 total strata. An additional stratum was subsequently added to the study when additional schools were added. Within each stratum, schools are assigned to the mentoring treatment, the information treatment, or the control group after sorting by a random number generator. Randomization took place in Bern, Switzerland in May of 2024.

2.3 Data Collection

Figure 2 shows the overall intervention and data collection timeline. In all schools, we planed to conduct a baseline survey before the interventions (wave 1), a midline survey immediately after the interventions at the end of the school year (wave 2), and an endline survey six months later, around the beginning of the new school year (wave 3). However, due to funding changes, the endline survey started a calendar year past when the intervention ended.

2.3.1 Baseline Data Collection

Baseline ran from April 19 to July 15 2024. This long window reflects a wide array of unanticipated implementation challenges including major flooding events, enumerators resigning and new hires needing to be retrained, and political economy challenges with negotiations between unions and the government. The baseline data collection phase concluded with a final sample of 231 schools and 2,357 students. Schools were not informed about their study status until the baseline survey was concluded.

The survey was conducted with all sixth graders attending schools during our visits of the respective schools and lasted about 30-40 minutes. It covered questions about (i) school dropout and educational aspirations (ii) perceived returns to education (iii)

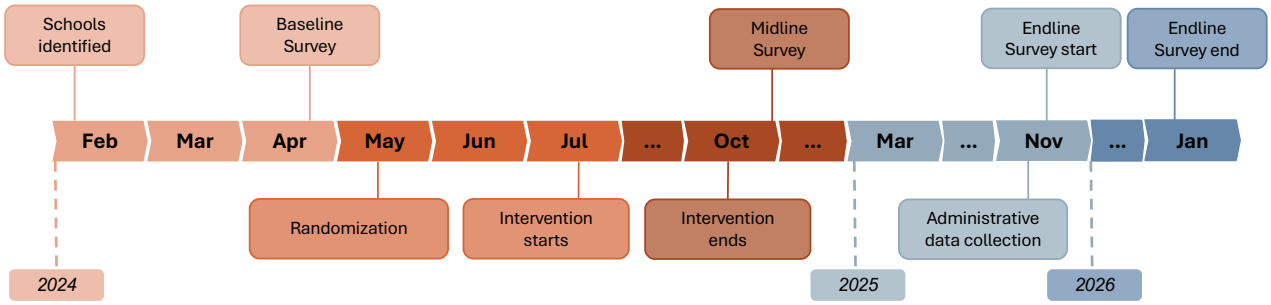


Figure 2: Timeline of data collection and intervention

child labor and allocation of the child's time (iv) migration and migrational intentions (v) awareness of child labor and human trafficking risks and (vi) socio-demographic background (only baseline). The baseline and midline surveys are conducted in school, whereas the endline survey is conducted in the child's home.

Innovation for Poverty Action (IPA)'s local data collection partner for baseline was Marketing Insight, who oversaw the recruitment process for the field team and enumerators. The enumerator role required people experienced in survey data collection. The baseline data collection utilized Survey CTO, as it ensures a secure and versatile platform. The data collection firm provided tablets to the field team. The collection instrument was pre-loaded onto the devices, allowing enumerators to conduct surveys offline. At the end of each day, the research team coordinated with the field team to ensure they correctly submitted all forms to a server. Enumerators conducted surveys in private, away from other students. Appendix 3 outlines the data quality efforts at baseline.

Marketing Insight mistakenly interviewed 2 schools at baseline that were not part of the study in replacement of 2 schools that were part of the study. This resulted in dropping these 2 schools and their 25 students (in total) from the baseline analysis but not from later surveys. We just lack baseline data for these 2 schools.

2.3.2 Midline Data Collection

The midline survey ran from October 16 to December 9, 2024. Midline data collection was timed to be complete before year-end exams begin in schools. IPA has employed a different data sub-contractor for midline data collection, Kantar, but similar systems were in place to assure the security and accuracy of data as in baseline. The baseline and midline surveys are very similar in content with a few minor changes in questions in the midline reflecting the experience of working with the baseline survey.

2.3.3 Endline Data Collection

The endline data started October 21, 2025 and ended on January 15, 2026.[4] IPA employed the baseline data collection partner, Marketing Insight, to conduct the endline survey. The endline is conducted in a home environment since children would have left the primary school they were enrolled in (to either enroll in a secondary school or because they drop out of school). Furthermore, surveying children in their homes allows for a broader range of questions, including some direct questioning of caregivers. We also conduct a Short Census form, which asks of community members whether subjects still live in the area, are in school, or are working. Overall, we have three main sources of information: student, caregiver and community members through the short census. Similar data quality efforts from baseline and midline are used.

2.3.4 Administrative Data

An earlier, partial version of the administrative records was provided to the PI team in May 2025 and analyzed in combination with the midline data. The analysis spelled out herein will be based on administrative data received by IPA in November 2025 and a second round expected in eary 2026 containing enrolled students in lower secondary and primary in the new academic year. This data allows tracking students that both enrolled in the next grade and that re-enrolled in the last year of primary. At the time of writing, there are still coding questions that need to be clarified that will effect whether we are able to use this data. This PI team has not analyzed this data at the time of writing. We believe this data includes information on whether students passed 6th grade in the year of the intervention. Neither IPA nor the PI team has received grades, but our Ministry of Education partners have promised to provide grades for 2023, 2024, and 2025 in early 2026.

2.3.5 Attrition from the Sample

This study was primarily designed to follow students identified at baseline. As such, we focus on attrition by those students in the baseline survey. We have several different data sources that feed into our culminating outcomes and intermediate outcomes. Because of the large, distinct volume of data sources, we have different attrition rates depending on the outcome variable. To keep the attrition discussion readable, we will compute attrition rates for enrollment in school in the 2025 academic year, child employment, migrated, the caregiver endline survey, the subject endline survey, and the midline survey. We anticipate an attrition rate of around 10 percent at endline. We will examine the impact of attrition on the internal validity of random assignment and on the comparability of the endline

[4]This revision to the PAP is being completed prior to receiving any of the endline data even though data collection ended a week prior.

subsample with the baseline subsample, as discussed in the section on balance below. If attrition invalidates the internal validity of random assignment, we anticipate following conventional approaches for bounding treatment effects to explore the sensitivity of our findings to different assumptions on the selection induced by attrition.

The midline survey was designed to be a snapshot of children attending school on the survey date. As such, there will be no recontact effort, and non-attendance (which is equivalent to attrition in the midline) is a primary outcome of interest.

To mitigate attrition at endline, our plan in our original PAP was to rely on teachers for help in locating their former students. However, because of tensions between our government partner and teacher's unions, we were unable to engage with teachers at endline. Recontact is based on the contact information students provided as at baseline and help from friends and neighbors. We suspect this will lead to more attrition than anticipated in the original PAP.

3 Empirical Analysis

3.1 Outcome Variables

The ultimate goal of the interventions evaluated in this study is to promote the transition from primary to lower secondary school. Accordingly, (1) school enrollment in lower secondary (also known as "Basic") constitutes the primary outcome of this study. Moreover, we will analyze effects on two related secondary outcomes that may also be affected by the interventions: (2) child employment and (3) migration.

Below, when we specify a "standardized index", we will apply the approach of Anderson (2008) to construct that index. When an index below is defined to include another standardized index, we create the (super) index using the variables that went into the input index. When creating these standardized indexes, we replace missing values with the control group mean. When looking at individual components of the indexes themselves, we do not replace missing values.

3.1.1 Culminating Outcomes

1. **Primary Outcome: Enrolled in School in the 2025 Academic Year**. We construct measures of both the mean and the mode across indicators from different sources of data at endline:

    - *Student survey:* Indicator that is one if participant reports they enrolled in either primary or lower secondary this academic year.

- *Caregiver survey:* Indicator that is one if caregiver reports the participant enrolled in either primary or lower secondary this academic year.
- *Short census form:* Indicator that is one if community member reports the participant enrolled in school in the current year.
- *Administrative data:* Indicator that is one if the participant appears as enrolled in either lower secondary or 6th grade of primary this academic year

In the case of having two modes, we first prioritize the participant's response, and then defer to administrative data second.

2. **Secondary Schooling Outcomes.** We construct various schooling outcomes out of data available at endline and administrative data:

   - **Passed Grade 6:** We construct measures of both the mean and the mode across indicators from different sources of data at endline:
     - *Student survey:* Indicator that is one if participant reports they passed their last year of primary in the previous school year.
     - *Caregiver survey:* Indicator that is one if the participant's caregiver reports they passed the last year of primary in the previous academic year.
     - *Administrative data:* Indicator that is one if the administrative data reports the participant passed their last year of primary in the previous academic year.

     In the case of having two modes, we first prioritize the participant's response, and then defer to administrative data second.
   - **Likes school:** Indicator that is one if student agrees that they like school (or liked it when they attended for dropouts.) (I like (liked) going to school)
   - **Currently attending school:** We construct measures of both the mean and the mode across indicators from different sources of data at endline:
     - *Student survey:* Indicator that is one if the participant has enrolled and will go to school again this year or has not missed school in the past week.
     - *Caregiver survey:* Indicator that is one if the caregiver reports the participant has enrolled and will go again to school this year or has not missed school in the last week.
     - *Short census form:* Indicator that is one if the community member reports the participant has gone to school recently.
     - *Administrative data:* Indicator that is one if the participant appears in admin data as enrolled and does not appear as having withdrawn.

In the case of having two modes, we first prioritize the participant's response, and then defer to administrative data.

- **Passed Grade 6, currently attending:** We construct measures of both the mean and the mode across indicators from different sources of data at endline:
  - *Student survey:* Indicator that is one if based on the student survey, the participant was coded as having advanced to lower secondary and is currently attending school.
  - *Caregiver survey:* Indicator that is one if based on the caregiver survey, the participant was coded as having advanced to lower secondary and is currently attending school.
  - *Administrative data:* Indicator that is one if based on the administrative data, the participant was coded as having advanced to lower secondary and is currently attending school.

  In the case of having two modes, we first prioritize the participant's response, and then defer to administrative data.
- **Not missed school in the past week:** We construct measures of both the mean and the mode across indicators from different sources of data at endline:
  - *Student survey:* Indicator that is one if the participant reports they have attended school this academic year and have not missed school in the past week.
  - *Caregiver survey:* Indicator that is one if the participant's caregiver reports they have attended school sometime this academic year and have not missed school in the past week.

  In the case of having two modes, we first prioritize the participant's response.

- **Passed first year of lower secondary:** Indicator that is one if student passed first year of lower secondary in Administrative data. The indicator is missing if final passing status is not available for a student.
- **Grades:** We are supposed to receive administrative records of year end grades for 2023, 2024, and 2025 by school subject. At the time of writing, we have not received these records and do not have a strong sense as to their form, completeness or usefulness. Hence, it is hard to pre-specify an approach to their analysis. We believe grades will be provided on a continuous scale. If our assumptions as to the nature of the data are correct, we will compute each of the following separately for each year:

- *Subject k Grade, year t* - Standardize grade in subject k in year t for school v. Standardization will be at the grade level, separately for each year t
- *Cumulative Grade, year t* - The simple mean of all available standardized subject grades in year t

3. Secondary Outcome: **Child Employment:** We construct measures of both the mean and mode across indicators from different sources at endline:

   - *Student survey:* Indicator that is one if the participant reports they have a job, as in something they do to earn money or help their family.
   - *Caregiver survey:* Indicator that is one if the caregiver reports the participant is has a job, as in something they do to earn money or help their family.
   - *Short census form:* Indicator that is one if the community member reports the participant has a job, as in something they do to earn money or help their family.

   In the case of having two modes, we first prioritize the student's response, and then defer to caregiver's responses.

4. Secondary Outcome: **Migration:** We construct migration outcomes from endline and administrative data:

   - **Migrated:** We construct measures of both the mean and mode across indicators from different sources at endline:
     - *Student survey:* Indicator that is one if the participant reports that during the week, they do not sleep in the same house they slept in around the same time last year.
     - *Caregiver survey:* Indicator that is one if the caregiver reports that during the week, the participant does not sleep in the same house the participant slept in around the same time last year.
     - *Short census form:* Indicator that is one if the community member reports that the participant's family no longer lives in the community. The indicator is also one if the family lives in the community, but the community member reports the participant does not spend most of their time in the community anymore.
     - *Administrative data:* Indicator that is one if the student appears enrolled in an area different than where they are found at baseline/midline.
     - *Field notes:* If the participant or their caregiver are not found for surveys, field team notes are used to determine whether participants were not found because they moved out of the community. If determined that the participant moved, this indicator is one.

In the case of having two modes, we first prioritize the student's response, then caregiver, then administrative data.

3.1.2 Explanation of Differences in Culminating Outcomes from Prior PAP

The 3 culminating outcomes have not changed. We have adjusted how we have defined each of these outcomes to reflect the data that is available to us after the endline survey. Our previous PAP did not anticipate the multiple sources of redundant information that we have now. Our change in funding for the endline also eliminated the need to be able to measure child labor as an outcome. Distinguishing between child labor and child employment is extremely costly in terms of interview complexity and time, and because we no longer had a specific deliverable due to our funder around measuring child labor, we refocused that outcome around the much simpler to measure child employment.

3.1.3 Intermediate Outcomes Measured at Endline

Intermediate outcomes are defined as those directly impacted by treatment. We believe these are the mechanisms that lead to our culminating outcomes. Other than attendance, endline components are only available for subjects with a in–person completed student or caregivers survey at endline (depending on the question). These intermediate outcomes are constructed only with data from the endline surveys.

- **Subject Beliefs:** We measure four different aspects of beliefs that might change with treatment. To reduce multiple hypothesis testing concerns, we will return each of these four belief measures, and will only report responses to component questions when they are informative to our speculation around mechanisms.
  - **Returns to Schooling** - An index based on:
    - Earn more: "People who complete lower secondary earn more money than those who only complete primary school."
    - Job requirement: "The jobs I want in the future will require that I complete at least lower secondary school."
    - Better life: "People in my community who completed lower secondary have a better life than those who did not."
  - **Pro-work - Schooling Tradeoffs** - An index based on:
    - Work useful:"I think I would learn more useful life skills by working now rather than going to school."
    - Prefer work: "If I could choose between going to school or working, I would prefer to work."

- **Must move for schooling returns** - Standardized response to:
  - "If people in my community want to get a better job after completing lower secondary school, they need to move to another place."
- **Migration Plans** - An index based on:
  - Will move: "I believe that someday I will move out of this community."
  - International plans: "I believe that someday I will live outside of Guatemala."

- **Capacity Constraints:** We think about capacity constraints as having six different dimensions. As with subject beliefs, we expect to report all six constructed aggregates and to only examine component questions when they are informative about underlying changes.
  - **Self-Regulation** - An index based on:
    - *Time Management*
      - Can find time: "If I make an effort, I can find enough time to study and do homework at home."
    - *Grit*
      - Worried of mistakes: "When I'm learning something new, I worry that others will see me make a mistake." Response order reversed so that higher values are less worried.
      - Motivated: "If something interests me, I will find a way to motivate myself and work hard on it."
  - **Accountability** - An index based on:
    - *Accountability and monitoring*
      - Someone cares: "There was someone outside my family who would feel disappointed if I stopped studying or learning."
      - Someone notice: "I believed that someone outside my family would notice if I stopped trying to learn new things or improve as a person."
    - *Salience*
      - "I talked with someone outside my family about how education is related to opportunities in the future."
  - **Planning** - An index based on:
    - *Knows how to continue school*: "I felt confused about how to continue with lower secondary education." Response order reversed so that higher values are less confused.

    - *Plans to enroll:* "I will go to school next year."
    - *Goal setting:*
        - Breaks down tasks: "When I have to do something difficult, I break it down into smaller steps, thinking about what to do first, second, and last"
        - Plans for goals: "When I think about my dreams and goals, I think about the different steps I need to take to achieve them"

- **Support** - An index based on:
    - *Parents engage with school*: "Did you or other adults who live in your household ever go to the school of the child in your care for any activity during the 2024 academic year?" This comes from the caregiver survey.
    - *Coping with shocks*: "When I have a problem, I have someone who is there for me."
    - *Home support*
        - Help at home: "If I had school homework and needed help, there would be someone in my home who could help me."
        - Education important:"My parents believe that education is important."
    - *Pro-school friends*
        - Friends value school: "My friends think that staying in school is important."
        - Friends go to school: "I have friends who go to school."
    - *Strength of Friendships*
        - More friends: " I believe I have more close friends now than a year ago."
        - Shares with friends: "I can talk to my friends about things that worry me"
    - *Community Support*
        - Someone Inspires me: "There was someone outside my family that inspired me to work harder to reach my goals."
        - Someone cares: "There was someone outside my family who cared if I succeeded at school or work."

- **Agency-** An index based on:
    - *Capable to decide:* "I feel capable of making my own decisions."
    - *Agency over time:* "I feel that my voice is heard and has weight in deciding what I do with my time."

- **Soft Skills:** The child's participation in a 30-minute survey with an adult stranger can be very informative about the child's soft skills, and we form an index based on 4 dimensions that we ask enumerators to assess.
  - *Attentive:* "Does the child pay attention during the survey and instructions?"
  - *Careful:* "Is the child careful and interested in the accuracy of the answers?"
  - *Alert:* "Is the child engaging and interactive?"
  - *Confident:* "Is the child shy or confident?"

- **Caregiver Beliefs**. We do not anticipate changes in caregiver beliefs, but we do anticipate readers having questions about whether they have changed. To inform this, we will examine the following 6 measures:
  - **Should complete basico:** "How much education do you think the children in this area should receive?" Transformed to "Thinks children in this area should complete at least basico".
  - **Age for Work:** "What would be the appropriate age for children to start helping the family financially, either with a job or by helping in the fields or family business?"
  - **Age to Stop School:** "At what age do you think children should stop going to school and pursue another full-time activity?"
  - **Age for Marriage:** "At what age do you think children should get married?"
  - **Expects Migration:** "Do you think the children in this community are likely to live or work outside of this area someday?"
  - **Age for Migration:** "What would be a good age for them to start doing it (work outside this area)?"

3.1.4 Intermediate Outcomes Measured at Midline Only

Because of changes in funders, our endline and midline surveys, 12 months apart and both post-intervention, contain different questions. At the time of revising this PAP, we believe the midline survey is primarily useful to understand difference in changes in beliefs between the mentoring and information arms. Because of data collection problems associated with the aforementioned teacher union - government problems, schools where we did not have a repeated relationship were non-cooperative with the midline. Non-random attrition plagues comparing the control group to either treatment arm. However, attrition looks random between the two treatment arms. Hence, in our final analysis, we expect the midline to be most useful in comparing the impact of the mentoring and information arms. The midline only information includes:

1. **Child labor knowledge** is a standardized index based on:
   - **Compulsory primary education:** Indicator that is one if respondent reports knowledge that it is not allowed to stop primary education to work in Guatemala.
   - **Compulsory lower secondary education:** Indicator that is one if respondent reports knowledge that it is not allowed to stop lower secondary education to work in Guatemala.
   - **Existence of restricted jobs:** Indicator that is one if respondent knows there are certain kinds of jobs you have to be 18 or older to perform.
   - **Identification of restricted jobs:** Fraction correctly identified by respondents among the following types of labor as restricted to those 18 or older:
     * Stone crushing
     * Construction
     * Street performers
     * Garbage recycling
     * Fireworks manufacturing
     * Domestic work in private homes
     * Repair of motorcycles and other vehicles

     At midline, the list of restricted jobs is as follows:
     * Stone crushing
     * Street performers
     * Garbage recycling
     * Fireworks manufacturing
     * Domestic work in private homes

   - **Identifies situations that are child labor**: Indicator that is one if respondent correctly identifies a given scenario as child labor. In the scenario, a student skips school to work until late at night.
   - **Identifies situations that are not child labor**: Indicator that is one if respondent correctly identifies a given scenario as not child labor. In the scenario, a student helps their family after school with house chores.
2. **Trafficking knowledge** is a standardized index based on:
   - **Heard of human trafficking:** Indicator that is one if respondent reports having heard of human trafficking before and correctly identifies what it is.

- **Human trafficking risk:** Indicator that is one if respondent reports both men and women, of any age, can be victims of human trafficking.
- **Domestic risk of human trafficking:** Indicator that is one if respondent reports that human trafficking can be a risk within Guatemala and does not require international travel.
- **Had a conversation about Human Trafficking:** Indicator that is one if respondent reports having a conversation about human trafficking in the past year.
- **Action on human trafficking situation:** Indicator that is one if respondent suggests a response to observing a human trafficking situation. Responses such as doing nothing or not knowing what to do as their only action if they recognized a situation of human trafficking are coded as 0.
- **Appropriate response to human trafficking:** Indicator that is one if respondent reports engaging in any of the following appropriate responses upon recognizing a situation of human trafficking:
  * Call the police
  * Talk to parents or family members
  * Talk to a teacher
  * Contact mayor or local authorities

- **Identifies human trafficking situations:** Indicator that is one if respondent correctly identifies a given scenario as being at risk of human trafficking. In the scenario, a stranger offers to take a student abroad to learn a new language, but offers no way to stay in touch.
- **Identifies a good decision if at risk of human trafficking:** Indicator that is one if respondent correctly identifies a given scenario as having a correct response to a potential human trafficking situation. In the scenario, a student worries about a friend who is getting a job offer too good to be true, and tells her teacher.
- **Identifies situations that increase risk of human trafficking:** Indicator that is one if respondent correctly categorizes all of the following situations on whether they increase or not the risk of being in danger of human trafficking:
  * Paying someone for a secret or unusual way to travel.
  * Someone promising you a job but not telling you what it is, where you will work, or how much you will earn.

* Traveling alone to another country.
* Telling your family and friends about your travel plans.
* Asking for help from trusted organizations or the government.

3. **Midline Well-being in school:** We measure psychological well-being using an adapted version of the EPOCH Measure of Adolescent Well-Being, which assesses 5 positive psychological characteristics (Engagement, Perseverance, Optimism, Connectedness, and Happiness) that might foster well-being, physical health, and other positive outcomes in adulthood (Kern et al., 2016). Each question asks respondents for a measure of "how true" the statement is on a scale of 1 to 5, where 5 is completely true. We create a standardized index based on the 10-item instrument answers. We use a 10-item instrument specifically adapted to in-school questioning of adolescents.[5]

   To understand and interpret these components of well-being, we follow the EPOCH recommendations for grouping of these 10 questions, and form the following indicators of these components of well-being. Don't know answers in these questions are replaced by control group means for each response. These are used as index components. Specifically:

   - **Engaged:** Indicator that is one if the respondent answers "completely true" or "very true" to both of these questions: "When I do an activity for school, I enjoy it so much that I lose track of time" and "I get so involved in my school assignments that I forget about everything else."
   - **Perseverant:** Indicator that is one if the respondent answers "completely true" or "very true" to both of these questions: "When I start a school assignment, I finish it" and "I am a person who works hard for school."
   - **Optimistic:** Indicator that is one if the respondent answers "completely true" or "very true" to both of these questions: "I am optimistic about my future in school" and "I believe that things in school will work out, no matter how difficult they seem."
   - **Connected:** Indicator that is one if the respondent answers "completely true" or "very true" to both of these questions: "When I have a problem, there is someone in school who is there for me" and "I have friends in school who are very important to me."

[5]This differs from the standard in the psych literature using EPOCH that just creates a simple average of the responses to the 10 questions (Buerger et al., 2023). We deviate to keep this consistent with our other indexes.

- **Happy:** Indicator that is one if the respondent answers "completely true" or "very true" to both of these questions: "I feel happy in school" and "I like going to school very much."

3.1.5 Treatment Saliency Measured at Endline

At endline, we collect data designed to measure whether subjects remember treatment. These are not independent of mechanisms above as many go into those mechanisms. We believe they also serve as a check on whether treatment remained salient at the time of endline. We will create a standardize index to summarize treatment saliency and examine individual responses for the following:

- School visits: "Did anyone visit your school during the last school year to hold sessions or provide information related to lower secondary school?"
- Inspiration: "There was someone outside my family that inspired me to work harder to reach my goals."
- Someone cares: "There was someone outside my family who cared if I succeeded at school or work."
- Salience: "I spoke to someone outside my family about how education is related to opportunities in the future"
- Someone cares about schooling:"There was someone outside my family that would feel disappointed if I stopped studying or learning."
- Someone notice: "I thought someone outside my family would notice if I stopped trying to learn new things or improve as a person."

3.2 Balancing Checks

3.2.1 Baseline Sample

Randomization took place in a controlled environment by the PI team. Hence, any differences in baseline characteristics across treatment arms are by chance. We evaluate balance across important background characteristics and primary outcomes at baseline. Specifically:

- Age in years at baseline
- Sex is female
- Spanish is spoken at home
- Household size
- Mother and father are both co-residents

- Household has running water
- Household has electricity
- Household has internet
- Student plans to enroll in lower secondary (basic)
- Student is in child labor
- Student plans to migrate internationally

The last three variables are chosen as baseline proxies for our three primary outcomes specified in our Theory of Change. The background characteristics from age - coresidency are demographic traits that we believe are relevant to our primary outcomes, and the three household characteristics are meant to proxy wealth.

We report means and standard deviations for each outcome variables denoted Y_{iv0} for outcome Y of individual i in school v at baseline (time 0) by treatment status. These summary statistics are in Table 3 after dropping the two schools that were interviewed in error by our baseline data collection firm. Our sample is 12.5 years old, balanced by gender, with a majority of families speaking Spanish at home. 70 percent have both parents present. Most households have running water and electricity with just over a third having internet access. Roughly two-thirds plan to enroll in lower-secondary, a majority are in child labor, and just under 15 percent plan to eventually migrate outside of Guatemala.

Overall, treatment status does not individually predict any characteristic or outcome variable shown in Table 3. The joint orthogonality chi-square is 18.8 (p-value=0.66), so we fail to reject the null that treatment assignment does not predict baseline characteristics. Furthermore, the bivariate orthogonality tests in the table show there are no significant differences in baseline characteristics when comparing the mentoring and control, information and control, and mentoring and information arms.

Alternatively, to expand the sample, we can replace missing values for baseline characteristics with midline values when baseline data is unavailable. This approach includes additional students who were interviewed at midline but missing from baseline. Summary statistics for this expanded sample are presented in Table 4. Baseline proxies for the three primary outcomes are not replaced using midline data, and thus are excluded from this table. There may be some meaningful differences in household size that merit attention in this sample.

We also conduct a series of hypothesis tests related to evaluating balance.

– For each outcome variable, we report the F-statistic and associated p-value for the null hypothesis that treatment status does not predict each individual outcome variable in the following regression:

$$Y_{iv} = \alpha + \gamma D_v + \lambda_1(A_i, F_i) + \lambda_2(S_v) + \upsilon_{iv} \tag{1}$$

  * γ is the vector of differences in Y_{iv0} associated with treatment status, and we report the F-statistic and p-value associated with the null that all of the components of the vector γ_0 is zero
  * $\lambda_1(A_i, F_i)$ is a vector of indicators for the child i's age and sex at baseline
  * $\lambda_2(S_v)$ is a vector of indicators for stratum
  * Standard errors allow for clustering by school in υ_{iv}

  The results of these F-tests on γ are reported in Column 4 of Table 3. We fail to reject the null of no difference in means associated with treatment for each of the variables in the table.

– We test the joint orthogonality of all the baseline outcome variable using Seemingly Unrelated Regressions by estimating (1) jointly across all the baseline characteristics listed above and reporting the Chi-Square test statistic and p-value associated with the null that the vector γ is zero across all equations.

  This Chi-Square test statistic is reported in Column 4 of Table 3 in the row labeled "Joint Orthogonality Chi-Square." The Chi-Square test statistic of 18 has an associated p-value of 0.66. We are not close to rejecting balance in these data.

– We also test the validity of randomization for comparing across study arms. For this test, we we evaluate the comparability of baseline attributes in these three comparisons:

  * Mentoring v. Control
  * Information v. Control
  * Mentoring v. Information

  To evaluate comparability of the first listed treatment group against the second listed treatment group, we limit the sample to these two groups, create an indicator that is one if the subject is in the first group, and regress this indicator on all the baseline variables plus the other components of the control function in (1) excluding age and sex:

$$D_{iv} = \beta_0 + \beta_1 Y_{iv}^1 ... + \beta_n Y_{iv}^n + \lambda_2(S_v) + \epsilon_{iv} \tag{2}$$

We exclude age and sex from the control function, because we want to evaluate compositional differences in age and sex in these tests. We report the F-statistics and p-value associated with the null hypothesis that all the baseline outcomes are jointly zero: $\beta_1 ... \beta_n = 0$. These results are in the last row of Table 3. In Column 1, we test the comparability of the mentoring arm to control at baseline. We have an F-stat of 0.92 with an associated p-value of 0.52. We are not close to rejecting the null of no difference in mentoring versus control. Column 2 compares the information arm to control. We have an F-stat of 0.5 with a p-value of 0.90. Column 4 compares mentoring to information, and we also fail to reject the null of no difference in background characteristics between those two treatment arms with an F-stat of 1.2 and an associated p-value of 0.32.

3.2.2 Attrition

While any differences in our balance tests at baseline will be by chance, attrition over time can invalidate the randomization. Our midline survey intends to capture snapshots of schools at the end of the school year and will have much higher student attrition than endline where our focus is on following up on baseline subjects.

There are two separate issues raised by attrition. Does attrition invalidate the internal validity of the treatment - control comparison? Does attrition imply that an internally valid comparison is no longer comparable to the original study comparison?

To evaluate the first questions of internal validity, we examine whether attrition is correlated with treatment status and we replicate the balance table described above around (1) on the recaptured sample.

To evaluate the second question of whether the resampled population is comparable to the baseline population, we compare (mean and standard deviation) the baseline characteristics of the original sample to the baseline characteristics of the recaptured sample. We then test with each characteristics differs with attrition in the same way we tested for randomization in 1:

$$Y_{iv} = \alpha + \gamma A_{iv} + \lambda_1(A_i, F_i) + \lambda_2(S_v) + v_{iv} \tag{3}$$

A_{iv} is an indicator that child i associated with primary school v was not recaptured in the follow-up survey round (midline or endline separately). We also test for joint significance with 2, using A_{iv} as the outcome rather than D_{iv}.

3.3 Treatment Effects

3.3.1 Intent to Treat

The basic analysis will be a regression of outcomes (Y) for individual i observed at time t in school v on an indicator for random assignment, a vector of age effects, a control for sex, and stratum fixed effects with standard errors clustered at the school level.

$$Y_{ivt} = \alpha_t + \gamma_t D_v + \lambda_{t1}(A_i, F_i) + \lambda_{t2}(S_v) + v_{ivt} \tag{4}$$

- We estimate (4) separately at the midline and endline time periods, reflected in the time-varying parameters in (4).
- D_v is a vector of indicators that indicate treatment assignment which takes place at the school v level
- $\lambda_{t1}(A_i, F_i)$ is a vector of indicators for the child i's age and sex at baseline
- $\lambda_{t2}(S_v)$ is a vector of indicators for stratum
- Standard errors v_{ivt} are clustered by school

3.3.2 Treatment Effects on the Treated

There are two types of non-compliance issues that we anticipate examining in our analysis.

- Not all schools randomly assigned to treatment implemented the program. There appears to be one school where, after the start of treatment, the principal refused to participate. There also appears to be a group of schools where our implementing partner failed to staff them appropriately. The exact scale of this problem is unclear at the time of writing. One definition of treated will be that the student was associated with a school where the program was implemented. We refer to this as associated with a **treated school**.
- Even in treated schools, not all students choose to attend our program. This might be because of disinterest in the program, a lack of consent, or absence from school. We define a **treated student** as a student who attends a large portion of the courses offered in the school. We define a large portion as having a session attendance rate of 92% or greater (the median attendance rate) for students in the mentoring arm. For the information arm, a treated student is defined as having attended all the treatment sessions offered in the school.

We anticipate adapting (4), replacing treatment assignment with ~~either treated school or~~ treated student and instrumenting for treated with treatment assignment in a two-stage least squares procedure.[6]

3.3.3 Analysis of Grades

Although we have not received data on grades at the time of writing, we anticipate receiving final grades for each subject in 2023, 2024, and 2025. If we receive these grades, we will analysis grades using the ITT and TOT as discussed above, but we will consider a secondary specification where we include the standardized 2023 grades for all available subjects as controls. We believe this specification will be important to isolate treatment effects on grades due to the wide variety of factors that influence grades.

3.3.4 2024 Class Composition Analysis

In addition to outcomes for the baseline study cohort, we examine whether the interventions induce compositional changes in Grade 6 enrollment during the treatment year. Because program delivery occurs at the school-by-grade level, all Grade 6 students present in treatment schools are exposed to the assigned intervention. As a result, treatment may draw students into the school and may change the composition of classrooms both through who enters and who stays.

We observe student-level year-end Grade 6 grades for all students enrolled in Grade 6 in each study school in 2024. Receipt of a year-end grade indicates completion of the academic year. Let $\mathcal{B}_v$ denote the set of Grade 6 students present at baseline in school v, and let $\mathcal{Y}_v$ denote the set of Grade 6 students in school v who receive a year-end grade in 2024. We define two distinct margins of compositional change.

Entry Margin. We define *entrants* as students who complete Grade 6 but were not present at baseline,

$$\mathcal{E}_v \equiv \mathcal{Y}_v \setminus \mathcal{B}_v.$$

Our primary entry outcomes are the number of entrants, $\mathcal{E}_v$, and the ratio of total Grade 6 completers to baseline Grade 6 enrollment, $\mathcal{Y}_v / \mathcal{B}_v$. These outcomes capture whether the interventions draw additional students into completing the academic year.

[6] We are not able to instrument treated school with random assignment as there is only one mentoring school that failed to implement the program, and this school was not interviewed at midline (students from this school were included in endline frame).

To characterize the type of students drawn in, we link all students to their Grade 5 outcomes from 2023 and compute the mean prior achievement of entrants,

$$\overline{Grade}_v^{2023,E} \equiv \frac{1}{|\mathcal{E}_v|} \sum_{i \in \mathcal{E}_v} Grade_{iv}^{2023},$$

where A_{iv}^{2023} denotes within school standardized Grade 5 achievement in 2023. We compare this to the mean prior achievement of baseline students who complete Grade 6 in 2024,

$$\overline{Grade}_v^{2023,S} \equiv \frac{1}{|\mathcal{S}_v|} \sum_{i \in \mathcal{S}_v} Grade_{iv}^{2023},$$

where $\mathcal{S}_v \equiv \mathcal{Y}_v \cap \mathcal{B}_v$. The difference $\overline{Grade}_v^{2023,E} - \overline{Grade}_v^{2023,S}$ summarizes how the academic profile of entrants compares to that of baseline students who persist.

Retention Margin. We also study retention among baseline students. Let $\mathcal{A}_v \equiv \mathcal{B}_v \setminus \mathcal{Y}_v$ denote baseline students who do not complete Grade 6. Our primary retention outcome is the baseline survival rate, $\mathcal{S}_v/\mathcal{B}_v$. To characterize selection into survival, we compare the mean prior achievement of baseline survivors and baseline attriters using Grade 5 outcomes from 2023. Differences in these means indicate whether the interventions change which baseline students persist to the end of the academic year.

Performance during the Treatment Year. To assess academic performance during Grade 6, we compute standardized Grade 6 outcomes in 2024 separately for entrants and baseline survivors,

$$\overline{Grade}_v^{2024,E} \quad \text{and} \quad \overline{Grade}_v^{2024,S},$$

where all grades are standardized within school prior to analysis. Differences between these outcomes describe how students drawn in by the interventions perform relative to baseline students who persist, recognizing that these comparisons reflect both performance conditional on completion and compositional change.

Estimation. The school level outcomes described above are:

- $\mathcal{E}_v$ - Number of Entrants
- $(\mathcal{Y}_v/\mathcal{B}_v) - 1$ - Percentage growth in students
- $\mathcal{S}_v$ - Number of Baseline Students that Survive to Yearend
- $\mathcal{A}_v$ - Number of Attritors from Baseline Students

- $\mathcal{S}_v / \mathcal{B}_v$ - Baseline Survival Rate
- $\overline{Grade}_v^{2023,E}$ - Mean prior GPA of Entrants
- $\overline{Grade}_v^{2023,S}$ - Mean prior GPA of Survivors
- $\overline{Grade}_v^{2023,A}$ - Mean prior GPA of Attritors
- $\overline{Grade}_v^{2023,E} - \overline{Grade}_v^{2023,S}$ - Difference in average prior grades between entrants and survivors
- $\overline{Grade}_v^{2024,E}$ - Mean GPA of Entrants
- $\overline{Grade}_v^{2024,S}$ - Mean GPA of Survivors
- $\overline{Grade}_v^{2024,E} - \overline{Grade}_v^{2024,S}$ - Difference in average grades between entrants and survivors

For each of these school-level outcomes Y_v, we estimate intent-to-treat effects using the randomized assignment of schools to treatment arms:

$$Y_v = \beta D_v + \lambda(S_v) + \varepsilon_v \tag{5}$$

where D_v is a vector of indicators for assignment to the mentoring and information interventions (with control omitted), $\lambda(S_v)$ are stratum fixed effects, and inference is based on heteroskedasticity-robust standard errors. In the case where the average GPA cannot be computed because of a lack of entrants, survivors, or attritors, that school will be dropped from the grade analysis.

3.3.5 Persistence of school-level effects on subsequent cohort

We also examine whether assignment to the interventions generates lingering school-level effects after the directly treated cohort has moved on. Although program delivery in 2024 is confined to the treated Grade 6 cohort, the interventions may affect teachers, siblings and other peers, school climate, or administrative engagement in ways that persist and influence later cohorts. To study this possibility, we analyze outcomes in academic year 2025 for the next Grade 6 cohort in the same study schools.

Let $Enroll_v^{2025}$ denote total Grade 6 enrollment in school v in 2025, as observed in administrative records. We also define $Assess_v^{2025}$ as the share of enrolled Grade 6 students in school v in 2025 with a recorded year-end grade, a measure of completing the academic year. We define $Grade_v^{2025}$ as the average year, standardized yearend grade for school v. We estimate 5 with these three dependent variables.

Because we do not observe pre-treatment outcomes for the same grade and schools at the time of writing, these estimates rely on randomized assignment within strata. The resulting coefficients should be interpreted as reduced-form evidence on the persistence of school-level effects to a subsequent cohort, rather than as peer effects or direct treatment effects on the 2025 cohort, and may reflect changes in school practices, school climate, composition, or administrative engagement induced by the intervention in 2024.

3.3.6 Spillovers within class in 2025

If we find an impact of the intervention on enrollment in 2025 and if we have data to estimate transition probabilities from our primary schools to specific lower secondary schools prior to 2024, we will examine spillovers in the first year of lower secondary school. The direction of such spillovers is theoretically ambiguous. Increased enrollment of study students could generate peer effects for other students, but could also create congestion or resource constraints that adversely affect outcomes.

We expect to observe enrollment and grade outcomes in 2025 for the first year of lower secondary school for the universe of lower secondary schools. Our spillover analysis focuses on students who are not part of the study sample. We restrict attention to lower secondary schools that receive students from study primary schools, and we exclude study participants themselves from all spillover outcome measures. Outcomes are aggregated at the lower secondary school level and include (i) enrollment of non-study students in the first year of lower secondary school and (ii) average academic performance of non-study students, measured using standardized grades.

A key challenge is that realized exposure to treated study students in a given lower secondary school is endogenous, as it depends on post-treatment enrollment decisions that may be correlated with unobserved local factors. To address this concern, we construct predicted exposure measures that rely only on random assignment and pre-intervention enrollment patterns.

Let primary schools be indexed by v and lower secondary schools by s. Let D_v^T denote an indicator for whether primary school v was assigned to the mentoring M, information intervention I, and the study population as a whole P. Let N_v^0 denote the baseline number of study-eligible students in primary school v. Using administrative data from cohorts prior to academic year 2025, we estimate π_{vs}, the probability that a student from primary school v enrolls in lower secondary school s in the first year of lower secondary education.

We then construct predicted exposure measures for each lower secondary school s:

$$\widehat{E}_s^T = \sum_v \pi_{vs} N_v^0 D_v^T, T \sim \{M, I, P\}$$

where D_v^T denotes assignment to group T. These predicted exposure measures capture the expected number of study students from each treatment arm and the study overall enrolling in lower secondary school s, based on random assignment and pre-period feeder flows.

At the time of writing, we do not have data on pre-intervention enrollment patterns although we are hopeful. If we are unable to get that data, as an alternative, we collect data from students at baseline and teachers at midline on where students expect to attend lower secondary school. These two records are incomplete but taken together, we should be able to assign study primary schools to lower secondary schools and compute predicted exposures by adding up study arm assignments.

Let E_s^M, E_s^I, and E_s^P denote the realized number of study students from each arm M, I and in the study who enroll in lower secondary school s in 2025. We estimate spillover effects using instrumental variables regressions that instrument realized exposure with predicted exposure. Specifically, for lower secondary school–level outcomes Y_s, we estimate:

$$Y_s = \beta_M E_s^M + \beta_I E_s^I + \lambda(S_v) + \varepsilon_s. \tag{6}$$

where $\lambda(S_v)$ denotes fixed effects for the randomization strata of all feeder primary schools. When there are feeder primaries from different strata, we include all relevant strata effects.

This approach isolates variation in exposure induced by random assignment in the study primary schools, while holding fixed historical enrollment patterns into lower secondary schools. The identifying assumption is that, conditional on stratum fixed effects and pre-intervention feeder flows, predicted exposure is uncorrelated with unobserved shocks to lower secondary school outcomes in 2025.

Spillovers in enrollment and academic performance capture distinct mechanisms and are therefore interpreted differently. Enrollment spillovers among non-study students reflect changes in schooling participation that may arise through peer effects, shifts in perceived school quality, or general equilibrium responses to changes in cohort size. In contrast, grade spillovers capture effects on academic performance conditional on enrollment and may reflect peer composition effects, changes in instructional resources, or congestion in classrooms. Enrollment spillovers speak primarily to participation

margins, while grade spillovers provide evidence on potential peer or congestion effects among students who remain enrolled.

The findings from estimating 6 are interpreted as the impact of adding a student who is in the mentoring arm, β_M, or information arm, β_I. If we control for the number of study subjects in the school E_s^P ($\widehat{E}_s^P$ is the instrument), this changes the interpretation of each coefficient to the impact of switching a study subject from control to the mentoring (information) arms. As such, adding E_s^P as a control in 6 controls for congestion externalities that owe strictly to the enrollment effect.

3.4 Heterogeneous Effects

These will be examined for culminating outcomes

3.4.1 Variable Definitions

We intend to look at heterogeneity in treatment effects based on the following variables:

- **Individual characteristics**
  - **Female**: A child is classified as female if the enumerator observes and records their sex as female during the interview.
  - **Older child**: A child is classified as older if they are older than 13 in grade 6. 12 is the modal age for grade 6, but 11-13 are a possibility with standard progression.
  - **Dropout risk**: We create an index of baseline covariates that we believe correlate strongly with dropout risk and split the sample by the median of the resulting index. At the time of writing, we expect the following baseline covariates as predictive of dropout risk. We also will look at heterogeneity separately by each of the components. This analysis will only be conducted for students with completed baseline survey.
    - **Student plans to drop out**. Students who did not report continuing school past primary in the baseline survey
    - **Student had missed school in the past week** at baseline
    - **Student is in child labor** at baseline
    - **Student plans to migrate** outside of Guatemala in the baseline survey.
- **Household wealth**: We ask about household wealth at baseline through enumerating a small list of household assets (TV, computer/laptop, fridge, car, motorcycle, running water, electricity/light, internet). We create a standardized index based on this inventory and split the population at the median of the index.

- **School characteristics**
  - **High dropout school:** High dropout school is defined as a school with above median dropout rates based on the average of the dropout rate in 2021 and the average of 2013-21. 2021 is the most recent dropout record available from the Ministry of Education. Because these data are used in stratification, tests of significance in balance tables will not condition on stratum fixed effects.
  - **Small school:** Small school is defined as below median grade 6 enrollment in 2024 based on Ministry of Education records. Because these data are used in stratification, tests of significance in balance tables will not condition on stratum fixed effects.
  - **Remoteness:** A primary school is classified as remote if it does not have a nearby lower secondary school.

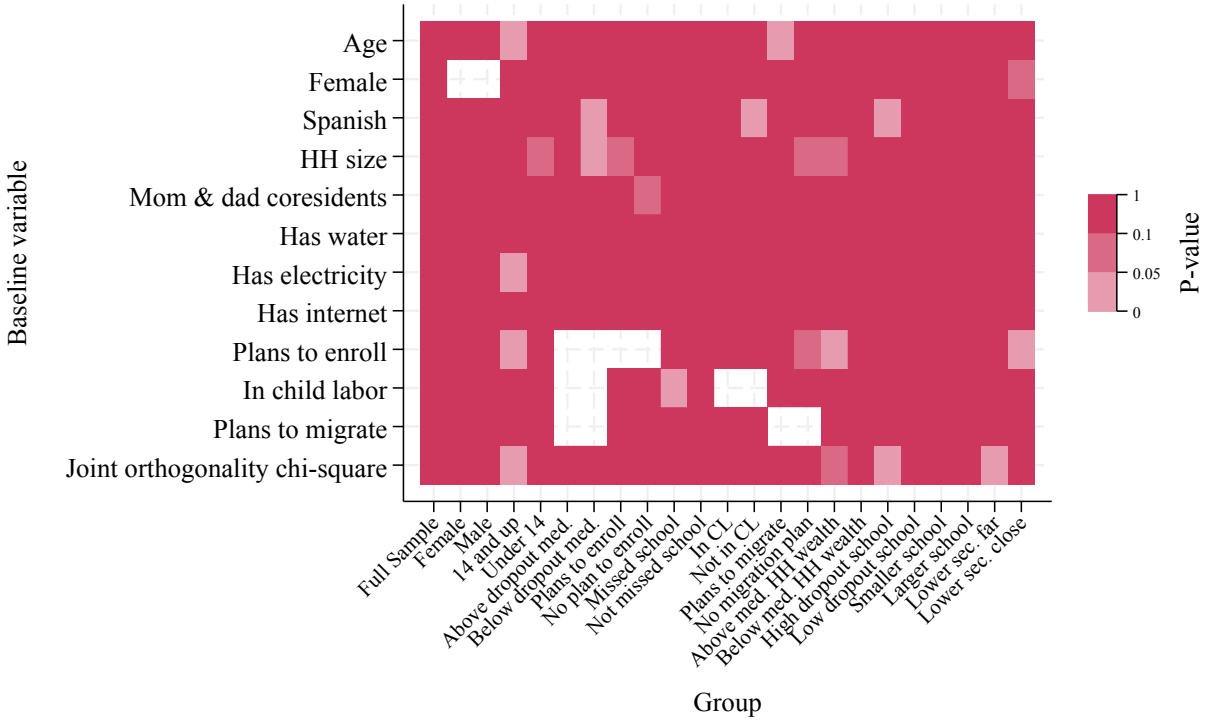
3.4.2 Balance Tests

Because randomization took place in controlled circumstances, we have the null hypothesis that it should be valid within subgroups. However, as some subgroups are small, we may reject this null in some subgroups if we happen to have imbalance across small groups. We evaluate the validity of randomization within subgroup using the same template as in the overall balance tests above.

Figure 3 summarizes the results of our analysis of balance for each form of heterogeneity that was proposed in the original PAP. A missing color indicates that the baseline characteristics is part of the heterogeneity. Hence, there is no within group variation to examine. The solid color indicates that we fail to reject the null hypothesis that random assignment do not predict the baseline variable. The lightest shade implies that differences in the row variable are statistically significant at conventional levels. We feel comfortable with comparisons of heterogeneity by sex and school size. Other sources of heterogeneity have some differences in baseline characteristics associated with random assignment that should be carefully considered when discussing differences. We discuss this further below.

- **Individual characteristics**
  - **Female**: Evaluations of the validity of randomization within female and male populations are in Tables 5 and 6 respectively. Based on the Joint-Orthogonality tests, we do have any concerns with the validity of the randomization for either sub-population.
  - **Older child**: Evaluations of the validity of randomization within older and younger populations are in Tables 7 and 8 respectively. Based on the

Figure 3: Balance within subgroups



Notes: This figure presents a heatmap of the p-values associated with each baseline variable, for the null hypothesis that treatment status does not predict each individual variable in Equation 1. The chi-square joint orthogonality tests the null that treatment assignment does not predict baseline characteristics.

Joint-Orthogonality tests, we are not concerned with using the randomization to study the younger population, but we will need to be attentive to covariates in our analysis for the older population. The older population is sufficiently small, that we end up with some differences in background characteristics across treatment arms.

* **Dropout Risk**: Because the baseline outcomes are included in the dropout index, they are omitted from the balance tests in Tables 9 and 10. Based on the joint orthogonality tests, we are not concerned about the validity of the randomization in either case even though there are some individual characteristics that merit attention (Spanish and household size in the low dropout risk group).
* **Student plans to drop out:** The data do not reject our hypothesis that randomization appears balanced for both students that plan to dropout and those that plan to enroll in lower secondary school in Tables 11 and 12.
* **Student had missed school in the past week** at baseline: Table 13 contains our findings for children that have missed school in the last week. We do not reject our joint null of no difference, although baseline child labor

merits attention. Table 14 contains our tests of our null of no difference with treatment for the group that has not missed school in the last week. We do not reject our null of no difference.

* **Student is in child labor** at baseline: Tables 15 and 16 have our findings for children that are and are not in child labor. We do not find any reason to reject our null, although there are some individual characteristics that merit attention (Spanish)
* **Student plans to migrate** outside of Guatemala at baseline: Tables 17 and 18 have our findings for children that do and do not eventually plan to migrate internationally at some point in their lives. We do not find any reason to reject our null. However there appears to be differences in age between the mentoring and information arms in the group that plans to migrate.
* **Household Wealth**: Tables 19 and 20 have our findings about the validity of randomization for households that are above and below median household wealth. For the below median wealth group in Table 20, we have no concerns about evaluating treatment effects within that sub-population. For wealthier households, we fail to reject orthogonality at 5 percent, but there may be some meaningful differences in plans for enrollment such that we should be attentive to the impact of inclusion of controls for those differences at baseline in our analysis.

– **School characteristics**

* **High dropout school:** We are concerned about evaluating treatment effects for schools that have high dropout rates and will have to be attentive to covariates. Our analysis of the validity of randomization is in Table 21. The data are more consistent with our null of no difference in the low dropout schools as evident in Table 22.
* **Small school:** The data are consistent with our null hypothesis of no difference with randomization for both small (Table 23) and large (Table 24) schools.
* **Remoteness:** The data are consistent with our null hypothesis of no difference with randomization for not remote schools (Table 25) , although we are mindful of differences in planning to enroll. For remote (Table 26) schools the sample is not balanced; we reject orthogonality at 5%.

3.4.3 Intent to Treat

Our analysis of heterogeneity in treatment effects will be limited to our intent-to-treat approach of (4). Specifically, for a given indicator H_{iv0} defined in the previous

subsection, we modify (4) as:

$$Y_{ivt} = \alpha_t^0 + \alpha_t^1 H_{iv0} + \gamma_t^0 D_v + \gamma_t^1 D_v * H_{iv0} + \lambda_{t1}(A_i, F_i) + \lambda_{t2}(S_v) + \upsilon_{ivt} \quad (7)$$

γ_t^0 is the impact of treatment assignment on Y_{ivt} when H_{iv0} is 0 and γ_t^1 tests the null that there is no difference in the impact of treatment assignment in the group where $H_{iv0} = 1$ compared to $H_{iv0} = 0$. $\gamma_t^0 + \gamma_t^1$ is how Y_{ivt} differs with treatment assignment compared to control.

For individual or school characteristics that are correlated, we also expect to conduct an analysis where H is a vector of characteristics, jointly estimating the interaction of treatment and the set of indicators.

3.5 Standard Error Adjustments

- Randomization takes place at the school level. All standard errors will be clustered by school.
- To address multiple hypothesis testing associated problems, outcomes are grouped into indexes as described in the data section above.
- In order to understand the meaning of indexes, individual components will be examined. To address concerns about multiple hypothesis testing within the components of indexes, we will report False Discovery Rate adjusted q-statistics in addition to conventional p-values (Anderson, 2008).

4 Research Team

- All PIs were involved in project development and design, survey instrument design, analysis of the data, and academic presentation of findings.
- Innovation for Poverty Action was the primary contractor for this project and was responsible for all contracting, sub-contracting, and award management. The IPA team included Sergio de Marco (Project Director), Laura Rodriguez (Research Manager for first two months of project), Daniel Hernandez Aldaco (Research Manager from month 3 through conclusion), Aurora Salvador Durand (Senior Research Associate until July 2023), Rosa Miranda Santa Cruz (Senior Research Associate from July 2023 on), Dalma Villanueva, and Victor Herencia.
- Prodessa implemented both programs. They were responsible for training all mentors and handling their employment.

- Marketing Insight was the data collection sub-contractor for IPA for the baseline and endline data collection. They oversaw the recruitment process for the field team and enumerators, supervised their work, and handled their employment.
- Kantar was the data collection sub-contractor for IPA for midline data collection. They oversaw the recruitment process for the field team and enumerators, supervised their work, and handled their employment.
- This project was funded in part as a cooperative agreement with the United States Department of Labor, and Tina Faulkner provided feedback on all aspects of this project.

5 Budget

This study is funded by a cooperative agreement between IPA and the U.S. Department of Labor for $1.5 million, a $200,000 award from the U.S. Department of State to Prodessa through IPA's HTRI initiative, and an additional $218,290 award from the U.S. Department of State to IPA for the endline survey .

6 Tables

Table 1: Differences between schools in the experiment and those not participating (in 2018)

	Other schools in the country (1)	Other schools in Quiché (2)	In the experiment (3)	Difference (3)-(2) (4)
% rural	90.98	95.12	100.00	4.88***
	(28.65)	(21.56)	(0.00)	(0.62)
	[15,419]	[1,229]	[231]	
% morning shift	94.57	97.72	98.70	0.98
	(22.67)	(14.93)	(11.35)	(0.86)
	[15,419]	[1,229]	[231]	
% bilingual	46.88	81.18	88.74	7.56***
	(49.90)	(39.10)	(31.67)	(2.36)
	[15,699]	[1,254]	[231]	
Total enrollment	135.95	118.80	115.02	-3.78
	(135.38)	(124.70)	(48.32)	(4.77)
	[15,363]	[1,226]	[231]	
Grd 6 enrollment	16.99	13.94	13.02	-0.92
	(19.34)	(16.03)	(6.00)	(0.60)
	[15,699]	[1,254]	[231]	
Number of teachers	5.68	4.90	4.57	-0.33*
	(4.95)	(4.63)	(1.91)	(0.18)
	[15,363]	[1,226]	[231]	
% with a school board	85.65	94.78	96.97	2.19*
	(35.06)	(22.24)	(17.18)	(1.30)
	[15,377]	[1,227]	[231]	

Notes: Column 1 displays the mean, standard deviation (in parentheses), and number of observations (in square brackets) for schools not in the experimental sample. Column 2 displays the mean, standard deviation (in parentheses), and number of observations (in square brackets) for schools not in the experimental sample, but in Quiché. Column 3 shows the mean, standard deviation (in parentheses), and number of observations (in square brackets) for schools in the sample. Column 4 reports the differences between the other schools in Quiché and those in the experimental sample, as well as the standard error of the difference (in parentheses). This data comes from Ministry of Education administrative records collected for Haimovich et al. (2021). Standard errors, clustered at the school level, are in parentheses. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

Table 2: Students enrolled in grade 6 in experimental vs. non-experimental schools

	Other schools in the country (1)	Other schools in Quiché (2)	In the experiment (3)	Difference (3)-(2) (4)
Panel A: 2018 academic year				
% male	51.26	52.03	52.69	0.66
	(49.98)	(49.96)	(49.94)	(1.05)
	[266,718]	[17,482]	[3,008]	
Age (Jan 1st, 2018)	12.83	13.01	13.14	0.13***
	(1.25)	(1.34)	(1.31)	(0.04)
	[264,439]	[17,338]	[2,998]	
GPA	7.61	7.40	7.35	-0.05
	(0.87)	(0.82)	(0.81)	(0.04)
	[260,656]	[17,279]	[2,984]	
% repeat grade	1.41	1.81	1.93	0.11
	(11.77)	(13.34)	(13.75)	(0.40)
	[266,661]	[17,480]	[3,008]	
% at-risk (statistical model)	41.44	66.55	86.89	20.33***
	(49.26)	(47.18)	(33.76)	(2.97)
	[248,695]	[16,229]	[2,837]	
% dropout	31.38	51.48	65.43	13.95***
	(46.40)	(49.98)	(47.57)	(2.02)
	[266,718]	[17,482]	[3,008]	
Panel B: 2021 academic year				
% male	51.10	52.14	50.91	-1.23
	(49.99)	(49.96)	(50.00)	(1.10)
	[282,955]	[18,349]	[3,204]	
% dropout	37.67	55.04	69.41	14.37***
	(48.46)	(49.75)	(46.08)	(1.96)
	[282,955]	[18,349]	[3,204]	

Notes: Column 1 displays the mean, standard deviation (in parentheses), and number of observations (in square brackets) for schools not in the experimental sample. Column 2 displays the mean, standard deviation (in parentheses), and number of observations (in square brackets) for schools not in the experimental sample, but in Quiché. Column 3 shows the mean, standard deviation (in parentheses), and number of observations (in square brackets) for schools in the sample. Column 4 reports the differences between the other schools in Quiché and those in the experimental sample, as well as the standard error of the difference (in parentheses). This data comes from Ministry of Education administrative records collected for Haimovich et al. (2021). Standard errors, clustered at the school level, are in parentheses. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

Table 3: Balance of Baseline Variables (Full sample)

	Mentoring Mean	Information Mean	Control Mean	F-Stat (p-value)
Variable	(1)	(2)	(3)	(4)
Age	12.544	12.527	12.553	0.051
	(1.287)	(1.189)	(1.383)	(0.950)
Female	0.496	0.492	0.485	0.125
	(0.500)	(0.500)	(0.500)	(0.882)
Spanish	0.652	0.581	0.643	1.293
	(0.477)	(0.494)	(0.479)	(0.277)
Household size	7.071	7.284	7.454	1.666
	(3.060)	(3.121)	(3.340)	(0.191)
Mom and dad are coresidents	0.715	0.694	0.698	0.374
	(0.452)	(0.461)	(0.459)	(0.688)
Household has running water	0.798	0.775	0.801	0.251
	(0.402)	(0.418)	(0.400)	(0.778)
Household has electricity	0.865	0.879	0.884	0.105
	(0.342)	(0.326)	(0.321)	(0.900)
Household has internet	0.373	0.347	0.372	0.436
	(0.484)	(0.476)	(0.484)	(0.647)
Plans to enroll in lower secondary	0.688	0.625	0.651	2.234
	(0.464)	(0.484)	(0.477)	(0.109)
Student is in child labor	0.558	0.583	0.583	0.555
	(0.497)	(0.493)	(0.493)	(0.575)
Plans to migrate outside Guatemala	0.143	0.134	0.147	0.120
	(0.350)	(0.341)	(0.354)	(0.887)
Joint Orthogonality Chi-Square				18.804
				(0.657)
Bivariate Orthogonality F-Stat	0.919 [a]	0.495 [b]		1.161 [c]
	(0.524)	(0.904)		(0.319)
Observations	762	761	808	

Notes: Columns 1-3 report the mean of each variable for the mentoring, information, and control groups, respectively, with standard deviations shown in parentheses. Column 4 reports the F-statistic and its associated p-value (in parentheses) for the null hypothesis that treatment status does not predict each individual outcome variable, controlling for age, gender, and stratum fixed effects, except for Age (does not include age fixed effects) and Female (does not include gender fixed effects). The chi-square joint orthogonality tests the null that treatment assignment does not predict baseline characteristics. [a] The bivariate test in Column 1 tests the hypothesis that there are no differences in baseline characteristics between the mentoring and control arms. [b] The bivariate test in Column 2 tests the hypothesis that there are no differences in baseline characteristics between the information and control arms. [c] The bivariate test in Column 4 tests the hypothesis that there are no differences in baseline characteristics between the mentoring and information arms. Orthogonality tests include stratum fixed effects. All standard errors are clustered at the school level.

Table 4: Balance of Baseline Variables (Full sample, replacing missing baseline data with midline responses)

	Mentoring Mean	Information Mean	Control Mean	F-Stat (p-value)
Variable	(1)	(2)	(3)	(4)
Age	12.544	12.527	12.553	0.051
	(1.287)	(1.189)	(1.383)	(0.950)
Female	0.496	0.492	0.485	0.125
	(0.500)	(0.500)	(0.500)	(0.882)
Spanish	0.638	0.576	0.636	1.128
	(0.481)	(0.495)	(0.482)	(0.326)
Household size	7.014	7.332	7.467	2.824
	(3.009)	(3.135)	(3.324)	(0.061)
Mom and dad are coresidents	0.705	0.689	0.698	0.202
	(0.456)	(0.463)	(0.459)	(0.817)
Household has running water	0.802	0.773	0.796	0.288
	(0.399)	(0.419)	(0.403)	(0.750)
Household has electricity	0.866	0.876	0.885	0.133
	(0.340)	(0.329)	(0.320)	(0.876)
Household has internet	0.372	0.356	0.371	0.212
	(0.484)	(0.479)	(0.483)	(0.809)
Joint Orthogonality Chi-Square				15.793
				(0.467)
Bivariate Orthogonality F-Stat	1.257 [a]	0.469 [b]		1.024 [c]
	(0.270)	(0.876)		(0.420)
Observations	853	841	859	

Notes: Columns 1-3 report the mean of each variable for the mentoring, information, and control groups, respectively, with standard deviations shown in parentheses. Column 4 reports the F-statistic and its associated p-value (in parentheses) for the null hypothesis that treatment status does not predict each individual outcome variable, controlling for age, gender, and stratum fixed effects, except for Age (does not include age fixed effects) and Female (does not include gender fixed effects). The chi-square joint orthogonality tests the null that treatment assignment does not predict baseline characteristics. [a] The bivariate test in Column 1 tests the hypothesis that there are no differences in baseline characteristics between the mentoring and control arms. [b] The bivariate test in Column 2 tests the hypothesis that there are no differences in baseline characteristics between the information and control arms. [c] The bivariate test in Column 4 tests the hypothesis that there are no differences in baseline characteristics between the mentoring and information arms. Orthogonality tests include stratum fixed effects. All standard errors are clustered at the school level.

Table 5: Balance of Baseline Variables (Female sample)

Variable	Mentoring Mean (1)	Information Mean (2)	Control Mean (3)	F-Stat (p-value) (4)
Age	12.437	12.445	12.437	0.006
	(1.164)	(1.128)	(1.199)	(0.994)
Spanish	0.644	0.562	0.656	1.851
	(0.480)	(0.497)	(0.476)	(0.160)
Household size	7.097	7.266	7.481	0.914
	(3.102)	(3.005)	(3.156)	(0.402)
Mom and dad coresidents	0.690	0.712	0.689	0.263
	(0.463)	(0.453)	(0.464)	(0.769)
Household has running water	0.774	0.758	0.787	0.180
	(0.419)	(0.429)	(0.410)	(0.835)
Household has electricity	0.849	0.887	0.881	0.634
	(0.359)	(0.317)	(0.324)	(0.532)
Household has internet	0.390	0.358	0.370	0.338
	(0.488)	(0.480)	(0.483)	(0.713)
Plans to enroll in lower secondary	0.692	0.624	0.641	1.949
	(0.462)	(0.485)	(0.480)	(0.145)
Student is in child labor	0.510	0.511	0.533	0.165
	(0.501)	(0.501)	(0.500)	(0.848)
Plans to migrate outside Guatemala	0.128	0.118	0.149	0.243
	(0.335)	(0.323)	(0.357)	(0.785)
Joint Orthogonality Chi-Square				16.299
				(0.698)
Bivariate Orthogonality F-Stat	0.538 [a]	0.639 [b]		1.133 [c]
	(0.861)	(0.779)		(0.342)
Observations	390	372	395	

Notes: Columns 1-3 report the mean of each variable for the mentoring, information, and control groups, respectively, with standard deviations shown in parentheses. Column 4 reports the F-statistic and its associated p-value (in parentheses) for the null hypothesis that treatment status does not predict each individual outcome variable, controlling for age, gender, and stratum fixed effects, except for Age (does not include age fixed effects). The chi-square joint orthogonality tests the null that treatment assignment does not predict baseline characteristics. [a] The bivariate test in Column 1 tests the hypothesis that there are no differences in baseline characteristics between the mentoring and control arms. [b] The bivariate test in Column 2 tests the hypothesis that there are no differences in baseline characteristics between the information and control arms. [c] The bivariate test in Column 4 tests the hypothesis that there are no differences in baseline characteristics between the mentoring and information arms. Orthogonality tests include stratum fixed effects. All standard errors are clustered at the school level.

Table 6: Balance of Baseline Variables (Male sample)

Variable	Mentoring Mean (1)	Information Mean (2)	Control Mean (3)	F-Stat (p-value) (4)
Age	12.650	12.607	12.662	0.127
	(1.390)	(1.241)	(1.529)	(0.880)
Spanish	0.661	0.599	0.630	0.606
	(0.474)	(0.491)	(0.483)	(0.547)
Household size	7.043	7.301	7.428	1.239
	(3.020)	(3.232)	(3.510)	(0.292)
Mom and dad coresidents	0.742	0.676	0.708	1.449
	(0.438)	(0.469)	(0.455)	(0.237)
Household has running water	0.823	0.792	0.814	0.214
	(0.383)	(0.407)	(0.390)	(0.807)
Household has electricity	0.882	0.871	0.886	0.088
	(0.323)	(0.335)	(0.318)	(0.916)
Household has internet	0.355	0.337	0.374	0.777
	(0.479)	(0.473)	(0.485)	(0.461)
Plans to enroll in lower secondary	0.683	0.627	0.662	1.007
	(0.466)	(0.484)	(0.474)	(0.367)
Student is in child labor	0.608	0.652	0.631	0.865
	(0.489)	(0.477)	(0.483)	(0.422)
Plans to migrate outside Guatemala	0.159	0.149	0.145	0.083
	(0.366)	(0.357)	(0.352)	(0.920)
Joint Orthogonality Chi-Square				18.913
				(0.528)
Bivariate Orthogonality F-Stat	0.734 [a]	0.682 [b]		1.183 [c]
	(0.692)	(0.740)		(0.306)
Observations	372	389	414	

Notes: Columns 1-3 report the mean of each variable for the mentoring, information, and control groups, respectively, with standard deviations shown in parentheses. Column 4 reports the F-statistic and its associated p-value (in parentheses) for the null hypothesis that treatment status does not predict each individual outcome variable, controlling for age, gender, and stratum fixed effects, except for Age (does not include age fixed effects). The chi-square joint orthogonality tests the null that treatment assignment does not predict baseline characteristics. [a] The bivariate test in Column 1 tests the hypothesis that there are no differences in baseline characteristics between the mentoring and control arms. [b] The bivariate test in Column 2 tests the hypothesis that there are no differences in baseline characteristics between the information and control arms. [c] The bivariate test in Column 4 tests the hypothesis that there are no differences in baseline characteristics between the mentoring and information arms. Orthogonality tests include stratum fixed effects. All standard errors are clustered at the school level.

Table 7: Balance of Baseline Variables (14 and older)

	Mentoring Mean	Information Mean	Control Mean	F-Stat (p-value)
Variable	(1)	(2)	(3)	(4)
Age	14.667	14.518	14.807	3.168
	(1.037)	(0.867)	(1.408)	(0.044)
Female	0.433	0.430	0.415	0.079
	(0.497)	(0.496)	(0.494)	(0.924)
Spanish	0.559	0.536	0.574	0.070
	(0.499)	(0.501)	(0.497)	(0.932)
Household size	7.824	7.509	7.347	0.376
	(3.539)	(3.067)	(3.667)	(0.688)
Mom and dad coresidents	0.706	0.618	0.693	2.154
	(0.458)	(0.488)	(0.464)	(0.120)
Household has running water	0.745	0.755	0.752	0.014
	(0.438)	(0.432)	(0.434)	(0.986)
Household has electricity	0.775	0.882	0.792	3.124
	(0.420)	(0.324)	(0.408)	(0.047)
Household has internet	0.333	0.309	0.366	1.265
	(0.474)	(0.464)	(0.484)	(0.285)
Plans to enroll in lower secondary	0.735	0.545	0.594	3.367
	(0.443)	(0.500)	(0.494)	(0.037)
Student is in child labor	0.686	0.694	0.640	0.406
	(0.466)	(0.463)	(0.482)	(0.667)
Plans to migrate outside Guatemala	0.118	0.155	0.149	0.864
	(0.324)	(0.363)	(0.357)	(0.424)
Joint Orthogonality Chi-Square				36.067
				(0.030)
Bivariate Orthogonality F-Stat	1.175 [a]	1.620 [b]		1.703 [c]
	(0.316)	(0.106)		(0.085)
Observations	102	110	101	

Notes: Columns 1-3 report the mean of each variable for the mentoring, information, and control groups, respectively, with standard deviations shown in parentheses. Column 4 reports the F-statistic and its associated p-value (in parentheses) for the null hypothesis that treatment status does not predict each individual outcome variable, controlling for age, gender, and stratum fixed effects, except for Age (does not include age fixed effects) and Female (does not include gender fixed effects). The chi-square joint orthogonality tests the null that treatment assignment does not predict baseline characteristics. [a] The bivariate test in Column 1 tests the hypothesis that there are no differences in baseline characteristics between the mentoring and control arms. [b] The bivariate test in Column 2 tests the hypothesis that there are no differences in baseline characteristics between the information and control arms. [c] The bivariate test in Column 4 tests the hypothesis that there are no differences in baseline characteristics between the mentoring and information arms. Orthogonality tests include stratum fixed effects. All standard errors are clustered at the school level.

Table 8: Balance of Baseline Variables (Under 14)

	Mentoring Mean	Information Mean	Control Mean	F-Stat (p-value)
Variable	(1)	(2)	(3)	(4)
Age	12.036	12.087	12.039	0.942
	(0.678)	(0.700)	(0.689)	(0.391)
Female	0.511	0.505	0.501	0.071
	(0.500)	(0.500)	(0.500)	(0.931)
Spanish	0.667	0.588	0.653	1.548
	(0.472)	(0.493)	(0.476)	(0.215)
Household size	6.955	7.246	7.469	2.720
	(2.965)	(3.131)	(3.293)	(0.068)
Mom and dad coresidents	0.717	0.707	0.699	0.243
	(0.451)	(0.456)	(0.459)	(0.784)
Household has running water	0.806	0.779	0.808	0.341
	(0.396)	(0.415)	(0.394)	(0.711)
Household has electricity	0.879	0.879	0.897	0.308
	(0.327)	(0.327)	(0.304)	(0.736)
Household has internet	0.379	0.353	0.373	0.344
	(0.485)	(0.478)	(0.484)	(0.709)
Plans to enroll in lower secondary	0.680	0.639	0.660	0.991
	(0.467)	(0.481)	(0.474)	(0.373)
Student is in child labor	0.538	0.564	0.575	0.762
	(0.499)	(0.496)	(0.495)	(0.468)
Plans to migrate outside Guatemala	0.147	0.131	0.147	0.270
	(0.354)	(0.337)	(0.354)	(0.764)
Joint Orthogonality Chi-Square				17.510
				(0.735)
Bivariate Orthogonality F-Stat	0.978 [a]	0.598 [b]		0.716 [c]
	(0.469)	(0.828)		(0.722)
Observations	660	651	708	

Notes: Columns 1-3 report the mean of each variable for the mentoring, information, and control groups, respectively, with standard deviations shown in parentheses. Column 4 reports the F-statistic and its associated p-value (in parentheses) for the null hypothesis that treatment status does not predict each individual outcome variable, controlling for age, gender, and stratum fixed effects, except for Age (does not include age fixed effects) and Female (does not include gender fixed effects). The chi-square joint orthogonality tests the null that treatment assignment does not predict baseline characteristics. [a] The bivariate test in Column 1 tests the hypothesis that there are no differences in baseline characteristics between the mentoring and control arms. [b] The bivariate test in Column 2 tests the hypothesis that there are no differences in baseline characteristics between the information and control arms. [c] The bivariate test in Column 4 tests the hypothesis that there are no differences in baseline characteristics between the mentoring and information arms. Orthogonality tests include stratum fixed effects. All standard errors are clustered at the school level.

Table 9: Balance of Baseline Variables (Above dropout index median)

	Mentoring Mean	Information Mean	Control Mean	F-Stat (p-value)
Variable	(1)	(2)	(3)	(4)
Age	12.388	12.509	12.350	1.334
	(1.060)	(1.141)	(1.057)	(0.265)
Female	0.501	0.465	0.490	0.742
	(0.501)	(0.499)	(0.500)	(0.477)
Spanish	0.541	0.539	0.607	1.974
	(0.499)	(0.499)	(0.489)	(0.141)
Household size	7.439	7.463	7.552	0.081
	(3.170)	(3.216)	(3.465)	(0.922)
Mom and dad coresidents	0.707	0.678	0.684	0.330
	(0.456)	(0.468)	(0.466)	(0.719)
Household has running water	0.779	0.737	0.792	0.986
	(0.415)	(0.441)	(0.406)	(0.375)
Household has electricity	0.872	0.882	0.879	0.004
	(0.334)	(0.323)	(0.326)	(0.996)
Household has internet	0.358	0.344	0.348	0.049
	(0.480)	(0.476)	(0.477)	(0.952)
Joint Orthogonality Chi-Square				12.938
				(0.677)
Bivariate Orthogonality F-Stat	0.513 [a]	0.867 [b]		0.833 [c]
	(0.845)	(0.546)		(0.575)
Observations	399	456	471	

Notes: Columns 1-3 report the mean of each variable for the mentoring, information, and control groups, respectively, with standard deviations shown in parentheses. Column 4 reports the F-statistic and its associated p-value (in parentheses) for the null hypothesis that treatment status does not predict each individual outcome variable, controlling for age, gender, and stratum fixed effects, except for Age (does not include age fixed effects) and Female (does not include gender fixed effects). The chi-square joint orthogonality tests the null that treatment assignment does not predict baseline characteristics. [a] The bivariate test in Column 1 tests the hypothesis that there are no differences in baseline characteristics between the mentoring and control arms. [b] The bivariate test in Column 2 tests the hypothesis that there are no differences in baseline characteristics between the information and control arms. [c] The bivariate test in Column 4 tests the hypothesis that there are no differences in baseline characteristics between the mentoring and information arms. Orthogonality tests include stratum fixed effects. All standard errors are clustered at the school level.

Table 10: Balance of Baseline Variables (Below dropout index median)

	Mentoring Mean	Information Mean	Control Mean	F-Stat (p-value)
Variable	(1)	(2)	(3)	(4)
Age	12.281	12.252	12.210	0.472
	(1.114)	(1.054)	(0.990)	(0.624)
Female	0.523	0.525	0.485	0.523
	(0.500)	(0.500)	(0.501)	(0.594)
Spanish	0.774	0.643	0.692	3.395
	(0.419)	(0.480)	(0.462)	(0.035)
Household size	6.667	7.016	7.317	3.673
	(2.885)	(2.957)	(3.157)	(0.027)
Mom and dad coresidents	0.725	0.718	0.719	0.152
	(0.447)	(0.451)	(0.450)	(0.859)
Household has running water	0.818	0.833	0.814	0.097
	(0.386)	(0.374)	(0.390)	(0.908)
Household has electricity	0.857	0.875	0.891	0.474
	(0.351)	(0.331)	(0.313)	(0.623)
Household has internet	0.388	0.351	0.405	0.919
	(0.488)	(0.478)	(0.492)	(0.401)
Joint Orthogonality Chi-Square				21.445
				(0.162)
Bivariate Orthogonality F-Stat	1.894 [a]	0.829 [b]		1.597 [c]
	(0.065)	(0.578)		(0.131)
Observations	363	305	338	

Notes: Columns 1-3 report the mean of each variable for the mentoring, information, and control groups, respectively, with standard deviations shown in parentheses. Column 4 reports the F-statistic and its associated p-value (in parentheses) for the null hypothesis that treatment status does not predict each individual outcome variable, controlling for age, gender, and stratum fixed effects, except for Age (does not include age fixed effects) and Female (does not include gender fixed effects). The chi-square joint orthogonality tests the null that treatment assignment does not predict baseline characteristics. [a] The bivariate test in Column 1 tests the hypothesis that there are no differences in baseline characteristics between the mentoring and control arms. [b] The bivariate test in Column 2 tests the hypothesis that there are no differences in baseline characteristics between the information and control arms. [c] The bivariate test in Column 4 tests the hypothesis that there are no differences in baseline characteristics between the mentoring and information arms. Orthogonality tests include stratum fixed effects. All standard errors are clustered at the school level.

Table 11: Balance of Baseline Variables (Plans to enroll)

	Mentoring Mean	Information Mean	Control Mean	F-Stat (p-value)
Variable	(1)	(2)	(3)	(4)
Age	12.309	12.355	12.243	0.901
	(1.130)	(1.071)	(1.003)	(0.408)
Female	0.515	0.487	0.480	0.339
	(0.500)	(0.500)	(0.500)	(0.713)
Spanish	0.714	0.620	0.668	1.752
	(0.452)	(0.486)	(0.471)	(0.176)
Household size	6.899	7.208	7.446	2.862
	(3.135)	(2.932)	(3.434)	(0.059)
Mom and dad coresidents	0.700	0.710	0.712	0.015
	(0.459)	(0.454)	(0.453)	(0.986)
Household has running water	0.800	0.807	0.814	0.046
	(0.401)	(0.395)	(0.389)	(0.955)
Household has electricity	0.859	0.874	0.886	0.414
	(0.349)	(0.332)	(0.318)	(0.661)
Household has internet	0.389	0.357	0.391	0.432
	(0.488)	(0.480)	(0.488)	(0.650)
Student is in child labor	0.577	0.586	0.593	0.320
	(0.495)	(0.493)	(0.492)	(0.726)
Plans to migrate outside Guatemala	0.134	0.139	0.144	0.097
	(0.341)	(0.346)	(0.352)	(0.908)
Joint Orthogonality Chi-Square				15.677
				(0.736)
Bivariate Orthogonality F-Stat	1.030 [a]	0.464 [b]		0.799 [c]
	(0.421)	(0.911)		(0.630)
Observations	524	476	527	

Notes: Columns 1-3 report the mean of each variable for the mentoring, information, and control groups, respectively, with standard deviations shown in parentheses. Column 4 reports the F-statistic and its associated p-value (in parentheses) for the null hypothesis that treatment status does not predict each individual outcome variable, controlling for age, gender, and stratum fixed effects, except for Age (does not include age fixed effects) and Female (does not include gender fixed effects). The chi-square joint orthogonality tests the null that treatment assignment does not predict baseline characteristics. [a] The bivariate test in Column 1 tests the hypothesis that there are no differences in baseline characteristics between the mentoring and control arms. [b] The bivariate test in Column 2 tests the hypothesis that there are no differences in baseline characteristics between the information and control arms. [c] The bivariate test in Column 4 tests the hypothesis that there are no differences in baseline characteristics between the mentoring and information arms. Orthogonality tests include stratum fixed effects. All standard errors are clustered at the school level.

Table 12: Balance of Baseline Variables (Plans to dropout)

	Mentoring Mean	Information Mean	Control Mean	F-Stat (p-value)
Variable	(1)	(2)	(3)	(4)
Age	12.399	12.491	12.383	0.408
	(0.983)	(1.177)	(1.078)	(0.666)
Female	0.504	0.491	0.504	0.307
	(0.501)	(0.501)	(0.501)	(0.736)
Spanish	0.517	0.516	0.596	2.325
	(0.501)	(0.501)	(0.492)	(0.100)
Household size	7.450	7.411	7.468	0.003
	(2.857)	(3.415)	(3.164)	(0.997)
Mom and dad coresidents	0.748	0.667	0.674	2.797
	(0.435)	(0.472)	(0.470)	(0.063)
Household has running water	0.794	0.723	0.777	0.618
	(0.405)	(0.448)	(0.417)	(0.540)
Household has electricity	0.878	0.888	0.879	0.119
	(0.328)	(0.316)	(0.326)	(0.888)
Household has internet	0.336	0.330	0.337	0.049
	(0.473)	(0.471)	(0.473)	(0.952)
Student is in child labor	0.517	0.578	0.566	0.765
	(0.501)	(0.495)	(0.497)	(0.467)
Plans to migrate outside Guatemala	0.164	0.126	0.152	0.783
	(0.371)	(0.333)	(0.360)	(0.459)
Joint Orthogonality Chi-Square				24.279
				(0.230)
Bivariate Orthogonality F-Stat	1.040 [a]	0.761 [b]		1.315 [c]
	(0.413)	(0.665)		(0.229)
Observations	238	285	282	

Notes: Columns 1-3 report the mean of each variable for the mentoring, information, and control groups, respectively, with standard deviations shown in parentheses. Column 4 reports the F-statistic and its associated p-value (in parentheses) for the null hypothesis that treatment status does not predict each individual outcome variable, controlling for age, gender, and stratum fixed effects, except for Age (does not include age fixed effects) and Female (does not include gender fixed effects). The chi-square joint orthogonality tests the null that treatment assignment does not predict baseline characteristics. [a] The bivariate test in Column 1 tests the hypothesis that there are no differences in baseline characteristics between the mentoring and control arms. [b] The bivariate test in Column 2 tests the hypothesis that there are no differences in baseline characteristics between the information and control arms. [c] The bivariate test in Column 4 tests the hypothesis that there are no differences in baseline characteristics between the mentoring and information arms. Orthogonality tests include stratum fixed effects. All standard errors are clustered at the school level.

Table 13: Balance of Baseline Variables (Has missed school in past week)

	Mentoring Mean	Information Mean	Control Mean	F-Stat (p-value)
Variable	(1)	(2)	(3)	(4)
Age	12.477	12.509	12.394	0.687
	(1.117)	(1.153)	(1.084)	(0.504)
Female	0.477	0.484	0.516	0.277
	(0.500)	(0.501)	(0.501)	(0.758)
Spanish	0.583	0.566	0.617	0.453
	(0.494)	(0.497)	(0.487)	(0.636)
Household size	7.364	7.801	7.986	1.921
	(3.307)	(3.282)	(3.706)	(0.149)
Mom and dad coresidents	0.701	0.687	0.708	0.202
	(0.459)	(0.465)	(0.456)	(0.817)
Household has running water	0.765	0.737	0.787	0.534
	(0.425)	(0.441)	(0.410)	(0.587)
Household has electricity	0.879	0.875	0.877	0.034
	(0.327)	(0.331)	(0.329)	(0.966)
Household has internet	0.352	0.391	0.368	0.472
	(0.479)	(0.489)	(0.483)	(0.624)
Plans to enroll in lower secondary	0.663	0.569	0.625	2.313
	(0.474)	(0.496)	(0.485)	(0.101)
Student is in child labor	0.672	0.770	0.766	3.117
	(0.470)	(0.422)	(0.424)	(0.046)
Plans to migrate outside Guatemala	0.193	0.135	0.144	2.114
	(0.396)	(0.343)	(0.352)	(0.123)
Joint Orthogonality Chi-Square				26.292
				(0.239)
Bivariate Orthogonality F-Stat	1.539 [a]	0.626 [b]		1.511 [c]
	(0.124)	(0.805)		(0.133)
Observations	264	281	277	

Notes: Columns 1-3 report the mean of each variable for the mentoring, information, and control groups, respectively, with standard deviations shown in parentheses. Column 4 reports the F-statistic and its associated p-value (in parentheses) for the null hypothesis that treatment status does not predict each individual outcome variable, controlling for age, gender, and stratum fixed effects, except for Age (does not include age fixed effects) and Female (does not include gender fixed effects). The chi-square joint orthogonality tests the null that treatment assignment does not predict baseline characteristics. [a] The bivariate test in Column 1 tests the hypothesis that there are no differences in baseline characteristics between the mentoring and control arms. [b] The bivariate test in Column 2 tests the hypothesis that there are no differences in baseline characteristics between the information and control arms. [c] The bivariate test in Column 4 tests the hypothesis that there are no differences in baseline characteristics between the mentoring and information arms. Orthogonality tests include stratum fixed effects. All standard errors are clustered at the school level.

Table 14: Balance of Baseline Variables (Has not missed school in past week)

	Mentoring Mean	Information Mean	Control Mean	F-Stat (p-value)
Variable	(1)	(2)	(3)	(4)
Age	12.263	12.346	12.239	0.921
	(1.064)	(1.086)	(1.000)	(0.400)
Female	0.530	0.492	0.474	1.602
	(0.500)	(0.500)	(0.500)	(0.204)
Spanish	0.689	0.590	0.656	1.566
	(0.463)	(0.492)	(0.475)	(0.211)
Household size	6.916	6.981	7.177	0.603
	(2.912)	(2.985)	(3.101)	(0.548)
Mom and dad coresidents	0.723	0.698	0.694	0.732
	(0.448)	(0.460)	(0.461)	(0.482)
Household has running water	0.815	0.798	0.808	0.109
	(0.388)	(0.402)	(0.394)	(0.897)
Household has electricity	0.857	0.881	0.887	0.272
	(0.350)	(0.324)	(0.317)	(0.762)
Household has internet	0.384	0.321	0.374	1.685
	(0.487)	(0.467)	(0.484)	(0.188)
Plans to enroll in lower secondary	0.701	0.658	0.665	1.060
	(0.458)	(0.475)	(0.472)	(0.348)
Student is in child labor	0.498	0.474	0.489	0.169
	(0.501)	(0.500)	(0.500)	(0.845)
Plans to migrate outside Guatemala	0.116	0.133	0.148	0.640
	(0.321)	(0.340)	(0.356)	(0.528)
Joint Orthogonality Chi-Square				20.840
				(0.531)
Bivariate Orthogonality F-Stat	1.009 [a]	0.616 [b]		1.469 [c]
	(0.441)	(0.813)		(0.149)
Observations	498	480	532	

Notes: Columns 1-3 report the mean of each variable for the mentoring, information, and control groups, respectively, with standard deviations shown in parentheses. Column 4 reports the F-statistic and its associated p-value (in parentheses) for the null hypothesis that treatment status does not predict each individual outcome variable, controlling for age, gender, and stratum fixed effects, except for Age (does not include age fixed effects) and Female (does not include gender fixed effects). The chi-square joint orthogonality tests the null that treatment assignment does not predict baseline characteristics. [a] The bivariate test in Column 1 tests the hypothesis that there are no differences in baseline characteristics between the mentoring and control arms. [b] The bivariate test in Column 2 tests the hypothesis that there are no differences in baseline characteristics between the information and control arms. [c] The bivariate test in Column 4 tests the hypothesis that there are no differences in baseline characteristics between the mentoring and information arms. Orthogonality tests include stratum fixed effects. All standard errors are clustered at the school level.

Table 15: Balance of Baseline Variables (Students in child labor)

	Mentoring Mean	Information Mean	Control Mean	F-Stat (p-value)
Variable	(1)	(2)	(3)	(4)
Age	12.423	12.475	12.343	1.052
	(1.168)	(1.088)	(1.058)	(0.351)
Female	0.468	0.430	0.446	0.465
	(0.500)	(0.496)	(0.498)	(0.629)
Spanish	0.622	0.595	0.620	0.103
	(0.486)	(0.491)	(0.486)	(0.903)
Household size	7.251	7.573	7.631	1.056
	(3.184)	(3.159)	(3.547)	(0.350)
Mom and dad coresidents	0.704	0.695	0.719	0.172
	(0.457)	(0.461)	(0.450)	(0.842)
Household has running water	0.811	0.768	0.806	0.561
	(0.392)	(0.422)	(0.396)	(0.571)
Household has electricity	0.872	0.882	0.874	0.076
	(0.334)	(0.323)	(0.332)	(0.926)
Household has internet	0.371	0.366	0.362	0.001
	(0.484)	(0.482)	(0.481)	(0.999)
Plans to enroll in lower secondary	0.709	0.630	0.661	1.972
	(0.455)	(0.483)	(0.474)	(0.141)
Plans to migrate outside Guatemala	0.170	0.145	0.139	0.727
	(0.376)	(0.353)	(0.346)	(0.485)
Joint Orthogonality Chi-Square				12.954
				(0.879)
Bivariate Orthogonality F-Stat	0.527 [a]	0.381 [b]		0.989 [c]
	(0.869)	(0.953)		(0.456)
Observations	423	440	469	

Notes: Columns 1-3 report the mean of each variable for the mentoring, information, and control groups, respectively, with standard deviations shown in parentheses. Column 4 reports the F-statistic and its associated p-value (in parentheses) for the null hypothesis that treatment status does not predict each individual outcome variable, controlling for age, gender, and stratum fixed effects, except for Age (does not include age fixed effects) and Female (does not include gender fixed effects). The chi-square joint orthogonality tests the null that treatment assignment does not predict baseline characteristics. [a] The bivariate test in Column 1 tests the hypothesis that there are no differences in baseline characteristics between the mentoring and control arms. [b] The bivariate test in Column 2 tests the hypothesis that there are no differences in baseline characteristics between the information and control arms. [c] The bivariate test in Column 4 tests the hypothesis that there are no differences in baseline characteristics between the mentoring and information arms. Orthogonality tests include stratum fixed effects. All standard errors are clustered at the school level.

Table 16: Balance of Baseline Variables (Students not in child labor)

	Mentoring Mean	Information Mean	Control Mean	F-Stat (p-value)
Variable	(1)	(2)	(3)	(4)
Age	12.239	12.283	12.218	0.187
	(0.965)	(1.100)	(0.984)	(0.829)
Female	0.567	0.575	0.546	0.358
	(0.496)	(0.495)	(0.499)	(0.700)
Spanish	0.690	0.565	0.678	3.964
	(0.463)	(0.497)	(0.468)	(0.020)
Household size	6.836	6.914	7.212	1.131
	(2.861)	(3.036)	(3.014)	(0.324)
Mom and dad coresidents	0.731	0.689	0.672	1.469
	(0.444)	(0.464)	(0.470)	(0.232)
Household has running water	0.779	0.790	0.797	0.059
	(0.415)	(0.408)	(0.403)	(0.942)
Household has electricity	0.854	0.886	0.896	0.457
	(0.354)	(0.319)	(0.306)	(0.634)
Household has internet	0.376	0.324	0.382	1.215
	(0.485)	(0.469)	(0.487)	(0.299)
Plans to enroll in lower secondary	0.657	0.622	0.636	0.995
	(0.476)	(0.486)	(0.482)	(0.371)
Plans to migrate outside Guatemala	0.110	0.121	0.155	0.745
	(0.314)	(0.326)	(0.363)	(0.476)
Joint Orthogonality Chi-Square				24.734
				(0.212)
Bivariate Orthogonality F-Stat	1.119 [a]	1.478 [b]		1.556 [c]
	(0.352)	(0.153)		(0.126)
Observations	335	315	335	

Notes: Columns 1-3 report the mean of each variable for the mentoring, information, and control groups, respectively, with standard deviations shown in parentheses. Column 4 reports the F-statistic and its associated p-value (in parentheses) for the null hypothesis that treatment status does not predict each individual outcome variable, controlling for age, gender, and stratum fixed effects, except for Age (does not include age fixed effects) and Female (does not include gender fixed effects). The chi-square joint orthogonality tests the null that treatment assignment does not predict baseline characteristics. [a] The bivariate test in Column 1 tests the hypothesis that there are no differences in baseline characteristics between the mentoring and control arms. [b] The bivariate test in Column 2 tests the hypothesis that there are no differences in baseline characteristics between the information and control arms. [c] The bivariate test in Column 4 tests the hypothesis that there are no differences in baseline characteristics between the mentoring and information arms. Orthogonality tests include stratum fixed effects. All standard errors are clustered at the school level.

Table 17: Balance of Baseline Variables (Plans to migrate)

	Mentoring Mean	Information Mean	Control Mean	F-Stat (p-value)
Variable	(1)	(2)	(3)	(4)
Age	12.211	12.578	12.252	5.322
	(0.924)	(1.066)	(0.985)	(0.006)
Female	0.459	0.431	0.496	0.191
	(0.501)	(0.498)	(0.502)	(0.826)
Spanish	0.532	0.480	0.605	0.823
	(0.501)	(0.502)	(0.491)	(0.441)
Household size	7.092	7.235	6.798	0.636
	(3.406)	(3.286)	(3.323)	(0.531)
Mom and dad coresidents	0.596	0.627	0.555	1.407
	(0.493)	(0.486)	(0.499)	(0.248)
Household has running water	0.807	0.804	0.832	0.302
	(0.396)	(0.399)	(0.376)	(0.740)
Household has electricity	0.844	0.882	0.908	1.153
	(0.364)	(0.324)	(0.291)	(0.319)
Household has internet	0.376	0.382	0.412	0.090
	(0.487)	(0.488)	(0.494)	(0.914)
Plans to enroll in lower secondary	0.642	0.647	0.639	0.178
	(0.482)	(0.480)	(0.482)	(0.837)
Student is in child labor	0.661	0.627	0.556	0.908
	(0.476)	(0.486)	(0.499)	(0.406)
Joint Orthogonality Chi-Square				27.288
				(0.127)
Bivariate Orthogonality F-Stat	0.801 [a]	1.352 [b]		1.840 [c]
	(0.628)	(0.216)		(0.063)
Observations	109	102	119	

Notes: Columns 1-3 report the mean of each variable for the mentoring, information, and control groups, respectively, with standard deviations shown in parentheses. Column 4 reports the F-statistic and its associated p-value (in parentheses) for the null hypothesis that treatment status does not predict each individual outcome variable, controlling for age, gender, and stratum fixed effects, except for Age (does not include age fixed effects) and Female (does not include gender fixed effects). The chi-square joint orthogonality tests the null that treatment assignment does not predict baseline characteristics. [a] The bivariate test in Column 1 tests the hypothesis that there are no differences in baseline characteristics between the mentoring and control arms. [b] The bivariate test in Column 2 tests the hypothesis that there are no differences in baseline characteristics between the information and control arms. [c] The bivariate test in Column 4 tests the hypothesis that there are no differences in baseline characteristics between the mentoring and information arms. Orthogonality tests include stratum fixed effects. All standard errors are clustered at the school level.

Table 18: Balance of Baseline Variables (Does not plan to migrate)

	Mentoring Mean	Information Mean	Control Mean	F-Stat (p-value)
Variable	(1)	(2)	(3)	(4)
Age	12.358	12.379	12.299	0.657
	(1.111)	(1.118)	(1.039)	(0.519)
Female	0.521	0.498	0.487	0.639
	(0.500)	(0.500)	(0.500)	(0.529)
Spanish	0.672	0.596	0.649	1.272
	(0.470)	(0.491)	(0.478)	(0.282)
Household size	7.067	7.291	7.567	2.833
	(3.001)	(3.097)	(3.332)	(0.061)
Mom and dad coresidents	0.735	0.704	0.723	0.740
	(0.442)	(0.457)	(0.448)	(0.478)
Household has running water	0.796	0.771	0.796	0.256
	(0.403)	(0.421)	(0.404)	(0.775)
Household has electricity	0.868	0.879	0.880	0.033
	(0.338)	(0.327)	(0.326)	(0.967)
Household has internet	0.372	0.341	0.365	0.471
	(0.484)	(0.475)	(0.482)	(0.625)
Plans to enroll in lower secondary	0.695	0.622	0.654	2.429
	(0.461)	(0.485)	(0.476)	(0.090)
Student is in child labor	0.541	0.576	0.588	1.165
	(0.499)	(0.495)	(0.493)	(0.314)
Joint Orthogonality Chi-Square				21.736
				(0.355)
Bivariate Orthogonality F-Stat	1.401 [a]	0.516 [b]		1.467 [c]
	(0.185)	(0.877)		(0.157)
Observations	653	659	690	

Notes: Columns 1-3 report the mean of each variable for the mentoring, information, and control groups, respectively, with standard deviations shown in parentheses. Column 4 reports the F-statistic and its associated p-value (in parentheses) for the null hypothesis that treatment status does not predict each individual outcome variable, controlling for age, gender, and stratum fixed effects, except for Age (does not include age fixed effects) and Female (does not include gender fixed effects). The chi-square joint orthogonality tests the null that treatment assignment does not predict baseline characteristics. [a] The bivariate test in Column 1 tests the hypothesis that there are no differences in baseline characteristics between the mentoring and control arms. [b] The bivariate test in Column 2 tests the hypothesis that there are no differences in baseline characteristics between the information and control arms. [c] The bivariate test in Column 4 tests the hypothesis that there are no differences in baseline characteristics between the mentoring and information arms. Orthogonality tests include stratum fixed effects. All standard errors are clustered at the school level.

Table 19: Balance of Baseline Variables (Above median household wealth index)

	Mentoring Mean	Information Mean	Control Mean	F-Stat (p-value)
Variable	(1)	(2)	(3)	(4)
Age	12.197	12.335	12.182	2.295
	(0.993)	(1.076)	(0.953)	(0.103)
Female	0.513	0.480	0.456	1.475
	(0.500)	(0.500)	(0.499)	(0.231)
Spanish	0.721	0.649	0.657	1.945
	(0.449)	(0.478)	(0.475)	(0.145)
Household size	6.973	7.589	7.662	2.641
	(3.192)	(3.629)	(3.481)	(0.074)
Mom and dad coresidents	0.729	0.703	0.710	0.274
	(0.445)	(0.458)	(0.454)	(0.760)
Household has running water	0.936	0.959	0.957	0.493
	(0.245)	(0.198)	(0.203)	(0.611)
Household has electricity	0.992	0.986	0.986	0.657
	(0.089)	(0.116)	(0.119)	(0.519)
Household has internet	0.628	0.561	0.559	1.308
	(0.484)	(0.497)	(0.497)	(0.273)
Plans to enroll in lower secondary	0.726	0.638	0.681	4.220
	(0.447)	(0.481)	(0.467)	(0.016)
Student is in child labor	0.560	0.584	0.589	0.272
	(0.497)	(0.494)	(0.493)	(0.762)
Plans to migrate outside Guatemala	0.130	0.144	0.165	0.733
	(0.337)	(0.352)	(0.372)	(0.482)
Joint Orthogonality Chi-Square				33.257
				(0.058)
Bivariate Orthogonality F-Stat	1.705 [a]	0.606 [b]		2.201 [c]
	(0.077)	(0.822)		(0.017)
Observations	376	367	417	

Notes: Columns 1-3 report the mean of each variable for the mentoring, information, and control groups, respectively, with standard deviations shown in parentheses. Column 4 reports the F-statistic and its associated p-value (in parentheses) for the null hypothesis that treatment status does not predict each individual outcome variable, controlling for age, gender, and stratum fixed effects, except for Age (does not include age fixed effects) and Female (does not include gender fixed effects). The chi-square joint orthogonality tests the null that treatment assignment does not predict baseline characteristics. [a] The bivariate test in Column 1 tests the hypothesis that there are no differences in baseline characteristics between the mentoring and control arms. [b] The bivariate test in Column 2 tests the hypothesis that there are no differences in baseline characteristics between the information and control arms. [c] The bivariate test in Column 4 tests the hypothesis that there are no differences in baseline characteristics between the mentoring and information arms. Orthogonality tests include stratum fixed effects. All standard errors are clustered at the school level.

Table 20: Balance of Baseline Variables (Below median household wealth index)

	Mentoring Mean	Information Mean	Control Mean	F-Stat (p-value)
Variable	(1)	(2)	(3)	(4)
Age	12.474	12.472	12.408	0.154
	(1.156)	(1.144)	(1.097)	(0.857)
Female	0.510	0.497	0.523	0.214
	(0.501)	(0.501)	(0.500)	(0.807)
Spanish	0.585	0.518	0.628	1.735
	(0.493)	(0.500)	(0.484)	(0.179)
Household size	7.166	7.000	7.232	0.523
	(2.927)	(2.531)	(3.173)	(0.593)
Mom and dad coresidents	0.702	0.685	0.686	0.111
	(0.458)	(0.465)	(0.465)	(0.895)
Household has running water	0.663	0.604	0.635	0.535
	(0.473)	(0.490)	(0.482)	(0.587)
Household has electricity	0.741	0.779	0.776	0.214
	(0.439)	(0.415)	(0.418)	(0.807)
Household has internet	0.124	0.147	0.173	2.148
	(0.330)	(0.355)	(0.379)	(0.119)
Plans to enroll in lower secondary	0.650	0.614	0.620	0.283
	(0.478)	(0.487)	(0.486)	(0.754)
Student is in child labor	0.556	0.582	0.577	0.421
	(0.497)	(0.494)	(0.495)	(0.657)
Plans to migrate outside Guatemala	0.155	0.124	0.128	1.078
	(0.363)	(0.330)	(0.334)	(0.342)
Joint Orthogonality Chi-Square				21.192
				(0.509)
Bivariate Orthogonality F-Stat	1.192 [a]	0.809 [b]		0.881 [c]
	(0.297)	(0.631)		(0.561)
Observations	386	394	392	

Notes: Columns 1-3 report the mean of each variable for the mentoring, information, and control groups, respectively, with standard deviations shown in parentheses. Column 4 reports the F-statistic and its associated p-value (in parentheses) for the null hypothesis that treatment status does not predict each individual outcome variable, controlling for age, gender, and stratum fixed effects, except for Age (does not include age fixed effects) and Female (does not include gender fixed effects). The chi-square joint orthogonality tests the null that treatment assignment does not predict baseline characteristics. [a] The bivariate test in Column 1 tests the hypothesis that there are no differences in baseline characteristics between the mentoring and control arms. [b] The bivariate test in Column 2 tests the hypothesis that there are no differences in baseline characteristics between the information and control arms. [c] The bivariate test in Column 4 tests the hypothesis that there are no differences in baseline characteristics between the mentoring and information arms. Orthogonality tests include stratum fixed effects. All standard errors are clustered at the school level.

Table 21: Balance of Baseline Variables (High dropout school)

Variable	Mentoring Mean (1)	Information Mean (2)	Control Mean (3)	F-Stat (p-value) (4)
Age	12.556	12.471	12.552	0.295
	(1.254)	(1.181)	(1.475)	(0.745)
Female	0.469	0.470	0.488	0.319
	(0.500)	(0.500)	(0.500)	(0.727)
Spanish	0.549	0.485	0.618	3.275
	(0.498)	(0.500)	(0.487)	(0.041)
Household size	7.218	7.252	7.808	2.076
	(2.885)	(3.079)	(3.616)	(0.130)
Mom and dad coresidents	0.742	0.701	0.675	1.437
	(0.438)	(0.458)	(0.469)	(0.242)
Household has running water	0.748	0.792	0.794	0.328
	(0.435)	(0.407)	(0.405)	(0.721)
Household has electricity	0.856	0.852	0.897	0.833
	(0.352)	(0.356)	(0.304)	(0.438)
Household has internet	0.399	0.318	0.382	1.939
	(0.490)	(0.466)	(0.487)	(0.149)
Plans to enroll in lower secondary	0.574	0.562	0.482	2.137
	(0.495)	(0.497)	(0.500)	(0.123)
Student is in child labor	0.560	0.632	0.590	1.531
	(0.497)	(0.483)	(0.493)	(0.221)
Plans to migrate outside Guatemala	0.160	0.121	0.163	0.918
	(0.367)	(0.326)	(0.370)	(0.402)
Joint Orthogonality Chi-Square				43.501
				(0.004)
Bivariate Orthogonality F-Stat	1.447 [a]	1.785 [b]		2.278 [c]
	(0.170)	(0.072)		(0.018)
Observations	326	365	369	

Notes: Columns 1-3 report the mean of each variable for the mentoring, information, and control groups, respectively, with standard deviations shown in parentheses. Column 4 reports the F-statistic and its associated p-value (in parentheses) for the null hypothesis that treatment status does not predict each individual outcome variable, controlling for age and gender, except for Age (does not include age fixed effects) and Female (does not include gender fixed effects). The chi-square joint orthogonality tests the null that treatment assignment does not predict baseline characteristics. [a] The bivariate test in Column 1 tests the hypothesis that there are no differences in baseline characteristics between the mentoring and control arms. [b] The bivariate test in Column 2 tests the hypothesis that there are no differences in baseline characteristics between the information and control arms. [c] The bivariate test in Column 4 tests the hypothesis that there are no differences in baseline characteristics between the mentoring and information arms. All standard errors are clustered at the school level. One school is missing the dropout indicator because it was not part of the original randomization sample but was visited at baseline and received treatment; no corresponding admin data is available for this school.

Table 22: Balance of Baseline Variables (Low dropout school)

Variable	Mentoring Mean (1)	Information Mean (2)	Control Mean (3)	F-Stat (p-value) (4)
Age	12.534	12.574	12.553	0.038
	(1.313)	(1.191)	(1.297)	(0.962)
Female	0.518	0.511	0.482	0.873
	(0.500)	(0.500)	(0.500)	(0.420)
Spanish	0.729	0.671	0.664	0.810
	(0.445)	(0.470)	(0.473)	(0.447)
Household size	6.961	7.321	7.157	0.546
	(3.183)	(3.184)	(3.063)	(0.581)
Mom and dad coresidents	0.695	0.689	0.718	0.383
	(0.461)	(0.464)	(0.450)	(0.683)
Household has running water	0.835	0.758	0.807	1.099
	(0.372)	(0.429)	(0.395)	(0.337)
Household has electricity	0.872	0.902	0.873	0.602
	(0.335)	(0.297)	(0.334)	(0.550)
Household has internet	0.353	0.373	0.364	0.107
	(0.479)	(0.484)	(0.482)	(0.899)
Plans to enroll in lower secondary	0.773	0.689	0.793	2.255
	(0.419)	(0.464)	(0.405)	(0.110)
Student is in child labor	0.557	0.539	0.578	0.345
	(0.497)	(0.499)	(0.494)	(0.709)
Plans to migrate outside Guatemala	0.131	0.149	0.134	0.261
	(0.337)	(0.357)	(0.341)	(0.771)
Joint Orthogonality Chi-Square				19.227
				(0.631)
Bivariate Orthogonality F-Stat	0.311 [a]	1.436 [b]		1.013 [c]
	(0.982)	(0.175)		(0.444)
Observations	436	389	440	

Notes: Columns 1-3 report the mean of each variable for the mentoring, information, and control groups, respectively, with standard deviations shown in parentheses. Column 4 reports the F-statistic and its associated p-value (in parentheses) for the null hypothesis that treatment status does not predict each individual outcome variable, controlling for age and gender, except for Age (does not include age fixed effects) and Female (does not include gender fixed effects). The chi-square joint orthogonality tests the null that treatment assignment does not predict baseline characteristics. [a] The bivariate test in Column 1 tests the hypothesis that there are no differences in baseline characteristics between the mentoring and control arms. [b] The bivariate test in Column 2 tests the hypothesis that there are no differences in baseline characteristics between the information and control arms. [c] The bivariate test in Column 4 tests the hypothesis that there are no differences in baseline characteristics between the mentoring and information arms. All standard errors are clustered at the school level. One school is missing the dropout indicator because it was not part of the original randomization sample but was visited at baseline and received treatment; no corresponding admin data is available for this school.

Table 23: Balance of Baseline Variables (Small school)

Variable	Mentoring Mean (1)	Information Mean (2)	Control Mean (3)	F-Stat (p-value) (4)
Age	12.475	12.471	12.457	0.016
	(1.371)	(1.230)	(1.217)	(0.984)
Female	0.496	0.505	0.447	0.713
	(0.501)	(0.501)	(0.498)	(0.493)
Spanish	0.656	0.619	0.632	0.110
	(0.476)	(0.487)	(0.484)	(0.896)
Household size	6.798	6.866	7.282	0.741
	(2.735)	(2.780)	(3.594)	(0.480)
Mom and dad coresidents	0.678	0.678	0.665	0.059
	(0.469)	(0.468)	(0.473)	(0.943)
Household has running water	0.792	0.822	0.852	0.776
	(0.407)	(0.384)	(0.356)	(0.463)
Household has electricity	0.923	0.847	0.842	2.031
	(0.267)	(0.361)	(0.366)	(0.137)
Household has internet	0.377	0.312	0.397	0.966
	(0.486)	(0.464)	(0.490)	(0.384)
Plans to enroll in lower secondary	0.617	0.624	0.694	1.130
	(0.487)	(0.486)	(0.462)	(0.327)
Student is in child labor	0.563	0.624	0.638	0.685
	(0.497)	(0.486)	(0.482)	(0.506)
Plans to migrate outside Guatemala	0.120	0.178	0.120	1.144
	(0.326)	(0.384)	(0.325)	(0.323)
Joint Orthogonality Chi-Square				23.301
				(0.385)
Bivariate Orthogonality F-Stat	1.633 [a]	1.316 [b]		0.661 [c]
	(0.112)	(0.236)		(0.769)
Observations	183	202	209	

Notes: Columns 1-3 report the mean of each variable for the mentoring, information, and control groups, respectively, with standard deviations shown in parentheses. Column 4 reports the F-statistic and its associated p-value (in parentheses) for the null hypothesis that treatment status does not predict each individual outcome variable, controlling for age and gender, except for Age (does not include age fixed effects) and Female (does not include gender fixed effects). The chi-square joint orthogonality tests the null that treatment assignment does not predict baseline characteristics. [a] The bivariate test in Column 1 tests the hypothesis that there are no differences in baseline characteristics between the mentoring and control arms. [b] The bivariate test in Column 2 tests the hypothesis that there are no differences in baseline characteristics between the information and control arms. [c] The bivariate test in Column 4 tests the hypothesis that there are no differences in baseline characteristics between the mentoring and information arms. All standard errors are clustered at the school level. One school is missing the school size indicator because it was not part of the original randomization sample but was visited at baseline and received treatment; no corresponding admin data is available for this school.

Table 24: Balance of Baseline Variables (Larger school)

Variable	Mentoring Mean (1)	Information Mean (2)	Control Mean (3)	F-Stat (p-value) (4)
Age	12.564	12.541	12.585	0.074
	(1.262)	(1.170)	(1.434)	(0.929)
Female	0.496	0.484	0.498	0.267
	(0.500)	(0.500)	(0.500)	(0.766)
Spanish	0.651	0.567	0.647	1.204
	(0.477)	(0.496)	(0.478)	(0.303)
Household size	7.157	7.442	7.513	1.359
	(3.153)	(3.239)	(3.248)	(0.261)
Mom and dad coresidents	0.727	0.701	0.710	0.325
	(0.446)	(0.458)	(0.454)	(0.723)
Household has running water	0.800	0.757	0.783	0.364
	(0.401)	(0.429)	(0.412)	(0.696)
Household has electricity	0.846	0.889	0.898	0.848
	(0.361)	(0.314)	(0.302)	(0.430)
Household has internet	0.371	0.359	0.363	0.042
	(0.484)	(0.480)	(0.481)	(0.959)
Plans to enroll in lower secondary	0.710	0.629	0.637	2.103
	(0.454)	(0.484)	(0.481)	(0.126)
Student is in child labor	0.557	0.570	0.564	0.073
	(0.497)	(0.496)	(0.496)	(0.929)
Plans to migrate outside Guatemala	0.150	0.120	0.157	1.035
	(0.358)	(0.325)	(0.364)	(0.358)
Joint Orthogonality Chi-Square				22.592
				(0.425)
Bivariate Orthogonality F-Stat	0.946 [a]	0.653 [b]		1.481 [c]
	(0.501)	(0.778)		(0.153)
Observations	579	552	600	

Notes: Columns 1-3 report the mean of each variable for the mentoring, information, and control groups, respectively, with standard deviations shown in parentheses. Column 4 reports the F-statistic and its associated p-value (in parentheses) for the null hypothesis that treatment status does not predict each individual outcome variable, controlling for age and gender, except for Age (does not include age fixed effects) and Female (does not include gender fixed effects). The chi-square joint orthogonality tests the null that treatment assignment does not predict baseline characteristics. [a] The bivariate test in Column 1 tests the hypothesis that there are no differences in baseline characteristics between the mentoring and control arms. [b] The bivariate test in Column 2 tests the hypothesis that there are no differences in baseline characteristics between the information and control arms. [c] The bivariate test in Column 4 tests the hypothesis that there are no differences in baseline characteristics between the mentoring and information arms. All standard errors are clustered at the school level. One school is missing the school size indicator because it was not part of the original randomization sample but was visited at baseline and received treatment; no corresponding admin data is available for this school.

Table 25: Balance of Baseline Variables (Lower Secondary Close)

	Mentoring Mean	Information Mean	Control Mean	F-Stat (p-value)
Variable	(1)	(2)	(3)	(4)
Age	12.443	12.565	12.533	0.183
	(1.318)	(1.242)	(1.235)	(0.833)
Female	0.495	0.532	0.457	2.621
	(0.501)	(0.500)	(0.499)	(0.079)
Spanish	0.658	0.619	0.651	0.217
	(0.476)	(0.486)	(0.477)	(0.805)
Household size	6.920	7.076	7.419	0.455
	(3.293)	(3.143)	(3.441)	(0.636)
Mom and dad coresidents	0.711	0.667	0.708	0.659
	(0.454)	(0.472)	(0.455)	(0.520)
Household has running water	0.760	0.768	0.737	0.352
	(0.428)	(0.423)	(0.441)	(0.705)
Household has electricity	0.933	0.911	0.870	0.956
	(0.250)	(0.285)	(0.337)	(0.389)
Household has internet	0.378	0.387	0.403	0.280
	(0.486)	(0.488)	(0.491)	(0.757)
Plans to enroll in lower secondary	0.693	0.635	0.733	3.389
	(0.462)	(0.482)	(0.443)	(0.039)
Student is in child labor	0.536	0.548	0.556	0.206
	(0.500)	(0.498)	(0.498)	(0.814)
Plans to migrate outside Guatemala	0.133	0.114	0.165	0.670
	(0.341)	(0.319)	(0.372)	(0.515)
Joint Orthogonality Chi-Square				30.186
				(0.114)
Bivariate Orthogonality F-Stat	1.300 [a]	1.250 [b]		0.800 [c]
	(0.254)	(0.278)		(0.640)
Observations	225	315	315	

Notes: Columns 1-3 report the mean of each variable for the mentoring, information, and control groups, respectively, with standard deviations shown in parentheses. Column 4 reports the F-statistic and its associated p-value (in parentheses) for the null hypothesis that treatment status does not predict each individual outcome variable, controlling for age, gender, and stratum fixed effects, except for Age (does not include age fixed effects) and Female (does not include gender fixed effects). The chi-square joint orthogonality tests the null that treatment assignment does not predict baseline characteristics. [a] The bivariate test in Column 1 tests the hypothesis that there are no differences in baseline characteristics between the mentoring and control arms. [b] The bivariate test in Column 2 tests the hypothesis that there are no differences in baseline characteristics between the information and control arms. [c] The bivariate test in Column 4 tests the hypothesis that there are no differences in baseline characteristics between the mentoring and information arms. Orthogonality tests include stratum fixed effects. All standard errors are clustered at the school level.

Table 26: Balance of Baseline Variables (Lower Secondary Distant)

	Mentoring Mean	Information Mean	Control Mean	F-Stat (p-value)
Variable	(1)	(2)	(3)	(4)
Age	12.581	12.502	12.565	0.403
	(1.274)	(1.153)	(1.463)	(0.669)
Female	0.497	0.465	0.501	1.460
	(0.500)	(0.499)	(0.500)	(0.236)
Spanish	0.650	0.554	0.638	1.443
	(0.477)	(0.498)	(0.481)	(0.239)
Household size	7.134	7.430	7.476	1.640
	(2.958)	(3.100)	(3.277)	(0.197)
Mom and dad coresidents	0.717	0.713	0.692	0.275
	(0.451)	(0.453)	(0.462)	(0.760)
Household has running water	0.814	0.780	0.842	1.230
	(0.390)	(0.415)	(0.365)	(0.295)
Household has electricity	0.836	0.857	0.893	1.239
	(0.371)	(0.351)	(0.310)	(0.292)
Household has internet	0.371	0.318	0.352	0.440
	(0.483)	(0.466)	(0.478)	(0.645)
Plans to enroll in lower secondary	0.685	0.619	0.599	2.136
	(0.465)	(0.486)	(0.491)	(0.122)
Student is in child labor	0.567	0.607	0.601	1.138
	(0.496)	(0.489)	(0.490)	(0.323)
Plans to migrate outside Guatemala	0.147	0.148	0.136	0.281
	(0.355)	(0.355)	(0.343)	(0.755)
Joint Orthogonality Chi-Square				35.057
				(0.038)
Bivariate Orthogonality F-Stat	2.645 [a]	0.669 [b]		1.779 [c]
	(0.005)	(0.764)		(0.067)
Observations	537	446	494	

Notes: Columns 1-3 report the mean of each variable for the mentoring, information, and control groups, respectively, with standard deviations shown in parentheses. Column 4 reports the F-statistic and its associated p-value (in parentheses) for the null hypothesis that treatment status does not predict each individual outcome variable, controlling for age, gender, and stratum fixed effects, except for Age (does not include age fixed effects) and Female (does not include gender fixed effects). The chi-square joint orthogonality tests the null that treatment assignment does not predict baseline characteristics. [a] The bivariate test in Column 1 tests the hypothesis that there are no differences in baseline characteristics between the mentoring and control arms. [b] The bivariate test in Column 2 tests the hypothesis that there are no differences in baseline characteristics between the information and control arms. [c] The bivariate test in Column 4 tests the hypothesis that there are no differences in baseline characteristics between the mentoring and information arms. Orthogonality tests include stratum fixed effects. All standard errors are clustered at the school level.

7 Appendix 1: Syllabus of Mentoring Sessions

7.1 Sessions with students

The following sessions are conducted in groups. After each group session, two individual 30-minute sessions are conducted. During these sessions, the mentor focuses on the particular situation of each student and provides personalized mentoring. Each student will have a minimum of two individual sessions as part of the program. The remaining individual sessions will be assigned to students who need them most as determined by the mentor.

7.1.1 Session 1: What do I want to be?

- Students become familiar with the program and the objective of the project
- Students reflect on their life plans and whether education will help them realize them

7.1.2 Session 2: Self-Esteem

- Children recognize that they have inner value and that they are important
- I can be wrong or make mistakes: Making mistakes is a normal part of life and does not make us bad or less valuable people
- Children recognize that they have the qualities to succeed in school and achieve their goals

7.1.3 Session 3: Getting to know professions and careers

- Children identify which professions and careers they could choose according to their tastes, preferences, goals, and dreams
- Children learn the routes of the educational system to become a professional

7.1.4 Session 4: Children's rights in the family context

- Explanation of children's rights: right to education, reproductive rights
- Discussion of the risks and consequences of dropping out of school
- Discussion of the risks, consequences, and prevention of teenage pregnancies

7.1.5 Session 5: Child Labor in Guatemala and Quiché

- Children learn about what is child labor and what is not, discussing consequences and risks
- Reflection on the distribution of each student's time in and out of school
- Life stages and priorities at each stage: Prioritizing school over work in childhood and adolescence

7.1.6 Session 6: Migration and Human Trafficking

- Children learn to identify types of migration and its causes
- Develop an understanding of independent migration and the risks faced by children and adolescents
- Definition and types of human trafficking
- Children learn about the frequent discourses used by traffickers and reflect on the risks faced by children and adolescents who become victims
- Learn about institutions and resources available for help

7.1.7 Session 7: Importance of lower secondary education for a better future

- Develop awareness of the impact that education in general, and lower secondary education in particular, can have on their lives
- Students recognize that learning is a beautiful thing that can help them grow and achieve their goals

7.1.8 Session 8: Getting to know my lower secondary school

- This session was actually conducted as the next to last session in all schools
- Mentors show students a video of a visit to a lower secondary school that is similar to a neighboring school (urban/rural)
- Mentors discuss with students how to enroll, attend, and transport themselves to the nearest lower secondary school

7.1.9 Session 9: Growth Mindset

- Children learn that skills and attitudes can be learned and developed with effort and practice, and are not fixed or determined.
- Children recognize that they can develop the qualities they still lack to succeed in school and in life

7.1.10 Session 10: Looking for alternatives to study

- Children learn alternatives to continue studying if their families' resources limit these possibilities
- Invite NGOs that provide scholarships or CCTs to inform about their programs

7.1.11 Session 11: Perseverance and effort

- Children understand the importance of completing homework assignments
- Children learn to be self-motivated to achieve their goals and do good quality tasks

Table 27: Number of Schools Experiencing Multiple Mentoring Sessions in One Day

Sessions Given in the Same Day	Number of schools
2 and 3	4
3 and 4	1
4 and 5	5
5 and 7	2
6 and 7	8
7 and 9	2
9 and 10	1
9, 10 and 11	1
9 and 11	1
10 and 11	7
13 and 14	1

These are sessions delivered in the same day. Each session is delivered in full.

7.1.12 Session 12: Making decisions

- Children understand the steps of decision-making and goal-setting, developing independence and initiative skills to make decisions.

7.1.13 Session 13: Communication and conflict resolution

- Children learn strategies for resolving conflicts or coping with problems
- Children can listen to and consider the opinions, views, and feelings of others

7.1.14 Session 14: My life project

- Children learn to differentiate between short-term and long-term goals
- Children make a life plan with their long-term goals
- Children define their short-term goals, which could help them realize their life project

7.1.15 Session Delivery Adjustments

Some schools experienced session delays due to local holidays or school activities. To ensure all schools in the mentoring arm received every session before the school year ended, sessions 12 and 13 were merged into a single session for 41 schools, covering the content of both sessions. Other schools received double sessions—two full sessions in one day—and only one school received three full sessions in a single day. The table below shows the number of schools receiving double sessions.

7.2 Sessions with caregivers

Work with caregivers involves three meetings with all caregivers, and individual meetings with prioritized caregivers in their home or a place of their choice (like a school) in case it was more convenient for them. A prioritized caregiver is the person in charge of a student with a higher risk of dropping out of school (low grades, lack of motivation, lack of parental support).

7.2.1 Initial meeting with caregivers

- Presentation of the program: issues, objectives, program activities aimed at teachers, caregivers, and students
- Describe role of caregivers in program development
- Basic information on returns to education, dropout rates, risks associated with child labor and human trafficking

7.2.2 Second meeting with caregivers

- Conducted after the first eleven weeks of mentoring with students
- Provide information on the transition to lower secondary
- Highlight benefits of education for the future of students, and the importance of lower secondary school for subsequent education opportunities
- Provide information on scholarships for different education levels
- Introduce locally famous people and their professions
- Discussion of progress made, challenges faced, proposals for the future and the role of caregivers in the continuity of the program
- Receive caregiver suggestions

7.2.3 Third meeting with caregivers

- Highlight children's rights, like access to education
- Discuss causes and consequences of school dropout, and provide information on returns to education and salary ranges by career and education level in Guatemala
- Describe child labor, its consequences, and how to report situations of child labor
- Discuss migration and risks related to unaccompanied child migration, including types of human trafficking and where to report potential human trafficking situations

7.2.4 Individual visits at home

- Talk with caregivers about the importance of education for the student's future. The student's presence was optional in these discussions
- Listen to caregiver suggestions and alternatives are proposed to continue with the student's education
- General tips for parents to support the student's education as provided by the mentor. Tailored advice was provided depending on the family/student situation
- At the end of the talk, the agreed commitments are highlighted

7.3 Sessions with teachers

7.3.1 Initial Session with teachers and director

- Presentation of the program: issues, objective, program activities aimed at teachers, caregivers and students
- Teacher suggestions for improving the program
- Discussion of each student's situation to focus on individual sessions
- Basic information on returns from education, school dropout rates, risks associated with child labor and human trafficking
- Role and commitment of teachers during the development of the program

7.3.2 Mid-term evaluation session

Note: This session was merged with the second caregiver session. Only the follow-up for each student detailed below happened only with teachers, without the parents' presence.

- Discussion of progress made, challenges faced, proposals for the future after the first eleven student sessions
- Role and commitment of teachers for program continuity
- Teacher feedback and suggestions
- Follow up of each student's progress and challenges (this was not part of the merged session with caregivers)

7.3.3 Final evaluation session

Note: This session was merged with the last caregiver session per teachers' request.

- Final presentation of program progress

- Role of teachers beyond the project, including agreements with teachers on how to promote the transition process to lower secondary education
- Final thanks to the teachers

7.4 Closing activity

After all sessions were completed, each school had a closing activity in which children, caregivers and teachers were awarded a diploma for completion of the sessions. Students were also gifted a student kit of school supplies (geometry tools, pencils, notebooks) and the caregivers were given a bag of groceries and supplies worth 20 Guatemalan Quetzales.

8 Appendix 2: Syllabus of Information Sessions

For the information treatment, two sessions were designed to align with the information covered in the mentoring sessions. Students received two separate sessions, while caregivers and teachers each had one session, followed by a joint session for both groups, totaling five scheduled sessions. Each session included a presentation followed by a discussion, with time for questions after each topic. The lack of computers and projectors in schools made it infeasible to have PowerPoint presentations. Instead, posters were printed to be explained and shown to the participants. At the end of each session, participants were given a leaflet summarizing the information covered.

8.1 Session 1

- Session content corresponds to student sessions 3, 7, 8, 10
- Leaflet includes information on returns to schooling, professions and educational opportunities, alternatives to study, etc.

8.2 Session 2

- Session content corresponds to student sessions 4, 5, 6, 7
- Leaflet includes information sheet on child labor, migration, human trafficking and returns to education (salaries).

8.3 Closing activity

After all sessions were completed, each school had a closing activity in which children, caregivers and teachers were awarded a diploma for completion of the sessions. Students were also gifted a student kit of school supplies (geometry tools, pencils,

notebooks) and the caregivers were given a bag of groceries and supplies worth 20 Guatemalan Quetzales.

9 Appendix 3: Monitoring strategies and data quality at Baseline

Several strategies were employed to assure data quality in the baseline survey.

- *WhatsApp chat groups:* First line of field monitoring and coordination, facilitating the communication between supervisors, enumerators, and the research team. Three separate groups were made:
  1. Enumerators' daily reports: The first group included field enumerators and the research team. The group's primary purpose was for enumerators to make daily productivity reports and notify the research team of any problems encountered during data collection. The research team used the group to make announcements of form updates and global feedback based on responses received during data collection.
  2. Supervisors' monitoring: The second group included the field supervisors and the research team. Its purpose was to inform supervisors of anything they should monitor closely with the enumerators, such as errors the research team could notice in the usage of the collection instrument by an enumerator or pending submissions from one of the teams.
  3. Coordination: The third group included the field coordinator, the data collection administrative team, and the research team. This group coordinated how the visits would be conducted during each week of the collection and inform of problems that could arise with the principals or locals around the schools that were visited.
- *High Frequency Checks (HFC):* HFCs are a series of programmed checks that the research team runs at least twice a day on the collected data. The HFCs are meant to ensure the quality of collected data and monitor closely the forms submitted to the server. The checks were useful in determining whether enumerators submitted all surveys each day, checking for duplicate submissions or errors, and checking the latest collection instrument was being used. The outputs obtained via the HFC were used to provide feedback to the enumerators and to contrast the productivity reported by the supervisors
- *Power BI dashboard:* To maintain all parties involved updated on the progress of the baseline collection, a Power BI dashboard was used to summarize the relevant

statistics, such as total number of schools surveyed and the average percentage of respondents. The dashboard also included information on the daily productivity of each enumerator, allowing supervisors to compare the information received by the database with the information reported by each enumerator.

– *Audits:* IPA programmed the surveys to record audio of certain sections to evaluate and ensure the correct application of survey protocols. IPA recorded the enumerator's application of some relevant questions in the sections of Child consent, Household, Education, Work, Migration, Human Trafficking and Recontact information. IPA audited a total of 200 surveys (8.5% of the total sample). The survey audit results indicated 92% were executed with no errors, 7% with few minor errors, and 1% had some audio problems that prevented the audit from being completed.

10 Appendix 4: Changes from previous PAP

The following outcome variables were removed due to funding changes:

– **Paid employment:** Respondent reports working for someone else for pay for one or more hours in the last week.

– **Employed**: Respondent reports working for someone else for pay, doing any kind of business activity, farming, or other activity to generate income, helping in a business or farm, for one or more hours in the last week.

– **Economically active:** Employed, temporarily absent from ongoing employment, participates in an unpaid apprenticeship, internship, or similar training, or helped in the production of goods for own consumption.

– **Unpaid household services:** Participated in unpaid household services in the last week.

– **Child labor**: Limited to children under 14. Employed unless they have a work permit, work 30 or fewer hours per week, and have not missed school in the last week.[7]

– **Child labor - EA:** Limited to children under 14. Economically Active unless they have a work permit, work 30 or fewer hours per week, and have not missed school in the last week. Guatemalan labor law emphasizes employment but standard

[7] Article 148 of Labor Code: Children under 14 cannot be employed. Article 149 of Labor Code: Ordinary work day can be diminished for persons under 18 to a maximum of 7 hours per day. Those under 14 with a permit can work a maximum of 6 hours per day. Article 150 of the Labor Code: Children under 14 can get a permit to work if in extreme poverty AND if they are meeting their required education by law.

child labor measures emphasize economic activity, and it is unclear whether the labor code intends to include the production of goods for own subsistence under prohibited employment.

- **Desires to migrate:** Indicator that is one if respondent reports they would want to live outside Guatemala as an adult if it were up to them. Children that have already migrated are coded as desiring to do so.
- **Ideal age to migrate:** Indicator that is one if respondent reports the ideal age to migrate is at 18 years or older.
- **Parents want to enroll:** Indicator that is one if respondent reports their parents want them to enroll in lower secondary in the coming year.

References

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.

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

Avitabile, C. and R. De Hoyos (2018). The heterogeneous effect of information on student performance: Evidence from a randomized control trial in mexico. *Journal of Development Economics 135*, 318–348.

Bettinger, E. P., B. T. Long, P. Oreopoulos, and L. Sanbonmatsu (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.

Buerger, S., J. Holzer, T. Yanagida, B. Schober, and C. Spiel (2023, June). Measuring Adolescents' Well-Being in Schools: The Adaptation and Translation of the EPOCH Measure of Adolescent Well-Being—A Validation Study. *School Mental Health 15*(2), 611–626.

Carrell, S. E. and M. Hoekstra (2014). Are school counselors an effective education input? *Economics Letters 125*(1), 66–69.

Dinkelman, T. and C. Martínez A (2014). Investing in schooling in chile: The role of information about financial aid for higher education. *Review of Economics and Statistics 96*(2), 244–257.

Edmonds, E., B. Feigenberg, and J. Leight (2023). Advancing the agency of adolescent girls. *Review of Economics and Statistics 105*(4), 852–866.

Haimovich, F., E. Vazquez, and M. Adelman (2021). Scalable early warning systems for school dropout prevention: Evidence from a 4.000-school randomized controlled trial. Technical report, Documento de Trabajo.

International Organization for Migration (IOM) and United Nations Population Fund (UNFPA) (2021). Characterization of International Migration in Guatemala (Census 2018).

Jensen, R. (2010). The (perceived) returns to education and the demand for schooling. *The Quarterly Journal of Economics 125*(2), 515–548.

Kern, M. L., L. Benson, E. A. Steinberg, and L. Steinberg (2016). The EPOCH Measure of Adolescent Well-Being. *Psychological Assessment 28*(5), 586–597.

Margaret Boittin, J. (2016). Reducing vulnerability to human trafficking: An experimental intervention using anti-trafficking campaigns to change knowledge, attitudes, beliefs, and practices in nepal.

McGuigan, M., S. McNally, and G. Wyness (2016). Student awareness of costs and benefits of educational decisions: Effects of an information campaign. *Journal of Human Capital 10*(4), 482–519.

Shah, M., S. Baird, J. Seager, B. Avuwadah, J. Hamory, S. Sabarwal, and A. Vyas (2024). Improving mental health of adolescent girls in low-and middle-income countries: Causal evidence from life skills programming. *Journal of Human Resources 59*(S), S317–S364.