Queen's Economics Department Working Paper No. 1472

Can Interventions Targeting Community Attitudes Improve Education for Marginalized Students? Evidence from a Mixed-Methods Experimental Design in Zimbabwe

Ardyn Nordstrom

Department of Economics
Queen's University
94 University Avenue
Kingston, Ontario, Canada
K7L 3N6

11-2021

Can Interventions Targeting Community Attitudes Improve Education for Marginalized Students? Evidence from a Mixed-Methods Experimental Design in Zimbabwe

Ardyn Nordstrom*

Job Market Paper - Current Version Here

Current version: November 19, 2021

Abstract. This paper uses a quasi-randomized field experiment in Zimbabwe to understand the impact of a large-scale intervention targeting community attitudes. I measure the impact that the program has had on attitudes, the behaviour of teachers and caregivers, and the learning and progression outcomes of at-risk youth. The quantitative survey and learning assessment data I use for this is complemented by transcripts from focus groups and interviews, which I analyze using innovative text mining methods to measure changes in community sentiment towards marginalized groups. I find that the program improved community attitudes toward girls' education by 0.403 SD over the three and a half year project. This contributed to a 20.9 percentage point increase in the likelihood that students in the treatment group reported receiving enough support from their community to continue learning during COVID-19 school closures, along with other changes in the behaviours of community members and families. The program facilitated better learning and progression outcomes, with marginalized students performing 0.28 SD better on learning assessments after the project. These findings lead to two important conclusions about the efficacy of interventions designed to reshape community attitudes. The first is that community attitudes can be influenced in a relatively short time to become more supportive towards marginalized groups. The second is that these interventions can support education outcomes. This paper also demonstrates the usefulness of qualitative methods and text mining techniques for future experimental work.

JEL Codes: I25, H43, C10
Keywords: Education, development, community attitudes, mixed-methods evaluation, quasi-randomized field experiment

*Queen's University, ardyn.nordstrom@queensu.ca.

I want to thank Christopher Cotton for excellent supervision. Additional thanks to Huw Lloyd-Ellis, Karen Ye, Bahman Kashi, Mbuso Jama, Gift Murombo, Janelle Zwier, Annah Matsika, Jose van Oosten, Shannon Veenstra, Jay MacKinnon, the broader evaluation teams at Jimat Development Consulting and Limestone Analytics, and the IGATE-T project team at World Vision. Additional thanks go to Rylen Sampson, who provided exceptional research assistance. This research was supported by financial support provided by the John Deutsch Institute for the Study of Economic Policy, the Ontario Graduate Scholarship, and the Social Sciences and Humanities Research Council of Canada. The paper has benefited from comments from seminar participants at several universities and seminars. Some of the analyses reported here contributed to the official endline evaluation for the IGATE-T project. All errors are my own.

1. Introduction

Parents, teachers, and community members all play an important role in shaping education outcomes. Parents and teachers establish the home and school environments where students develop and determine many of the opportunities available to students. At the same time, social and gender norms established at a community level contribute to a wide array of child and adolescent outcomes, including early marriage, the allocation of domestic and unpaid care work, and the education and employment opportunities available to women (Chang et al., 2020).[1] This paper uses a quasi-randomized field experiment to examine a large-scale intervention targeting the attitudes of community members towards marginalized adolescents—including girls and low-performing students—and measures the impact this has on community attitudes and education outcomes.

Significant work has been done to understand how gender gaps in education can be reduced through financial incentives (see Attanasio et al. (2012); Kremer et al. (2009); Bettinger (2012); Gneezy et al. (2011), among others), girls' empowerment campaigns (see Bandiera et al. (2020); Buchmann et al. (2017); Cotton et al. (2020)), improving interpersonal skills (Ashraf et al., 2020), investing in school infrastructure (see Adukia (2017); Barrera-Osorio et al. (2011); Birdthistle et al. (2011)), and mitigating early marriage (see Agarwal et al. (2021); Buchmann et al. (2017, 2021)). However, despite the importance of community attitudes in determining the opportunities available to girls and other marginalized groups, there is limited evidence surrounding the impact of programs that specifically target such attitudes at the individual or community level.

This paper is the first to explicitly examine the impact of an intervention targeting community-level attitudes. In a recent meta-analysis, Chang et al. (2020) highlight the need for additional research to understand the efficacy of interventions specifically targeting gender attitudes. Dhar et al. (2021) provides one of the only such studies by looking at the impact of an intervention designed to reshape the gender attitudes of adolescents. This paper differs from Dhar et al. (2021) in several key ways. The first is that, in contrast to the "Breakthrough" program examined in Dhar et al. (2021), the intervention studied in this paper not only targeted students' attitudes, but also

[1] Community-level factors also have significant impacts on child nutrition (Moestue and Huttly, 2008), and adolescent contraceptive use (Kravdal, 2002; Moursund and Kravdal, 2003), which both have consequences for education outcomes. A review of early marriage in developing countries Singh and Samara (1996) also highlights that girls from more rural areas are less likely to be exposed to modern values regarding marriage, which typically favour postponing marriage and allowing girls to have more autonomy over their relationships.

those of community members—including religious leaders, teachers, students' primary caregivers, and other parents in the community.[2] The second difference is that, in addition to considering the impact on attitudes themselves to show the efficacy of this type of intervention, this paper examines the impact that the program had on education outcomes, including progression. This paper also analyses attitudes by using text-mining techniques that show how implicit gender attitudes have changed.

To examine the impact of this intervention, I use a mixed-methods approach involving quantitative and qualitative evidence from a quasi-randomized field experiment. The main outcome is community attitudes (ie. what communities believe is "right" or "good" with respect to the education of girls and other marginalized students). Since the project is intended to ultimately improve education for marginalized students, the other main outcome of interest is education, which I evaluate by examining learning and progression. The quantitative evidence I use to evaluate these outcomes is based on impact measurements and heterogeneous treatment effects using survey panel data collected from control and treatment groups.

The qualitative evidence is based on findings from interviews and focus groups conducted in treatment areas that have been analyzed using sentiment analysis, a text mining technique that can measure attitudes toward sensitive topics (Xu et al., 2019; Liu et al., 2010).[3] To my knowledge, this is the first study to use sentiment analysis in an evaluation of a development intervention.

Incorporating qualitative data using rigorous text mining techniques provides two primary benefits. The first is that sentiment analysis offers a better way to measure implicit attitudes towards sensitive topics, such as gender norms, which can be difficult to measure (Nillesen et al., 2021). The second benefit is that the open-ended nature of qualitative interviews makes it possible to develop a more comprehensive understanding of the possible mechanisms contributing to changes in outcomes.

The project I evaluate was implemented by World Vision Zimbabwe, World Vision UK, and several local partner organizations between 2018 and 2021 in rural Zimbabwe.

[2]Throughout the paper, I refer to the individual who provides the primary care for a student at home as the student's "caregiver". In the majority of cases, this refers to the student's female parent. However, caregiver is a more accurate description since a significant proportion of the sample is orphaned or does not live with either parent.

[3]Mixed-methods approaches that incorporate evidence from qualitative data allow impact evaluations to answer a more diverse set of research questions. As highlighted by White (2013), randomized trials are the best tool to answer questions related to *attribution*, but mixed-methods approaches involving qualitative evidence are better suited to explaining the *mechanisms* that are relevant to the outcomes measured in a randomized design.

The "Improving Gender Attitudes, Transition, and Education" (IGATE-T) project operated in 318 schools in 238 communities, reaching over 120,000 students across rural Zimbabwe. The project took a multi-faceted approach to try to influence the attitudes that teachers, caregivers, local community members, and students held about gender, and to ultimately improve learning and progression outcomes for marginalized girls and boys in these communities.[4] Although the project specifically targeted girls, the project's emphasis on supporting marginalized students more broadly was also intended to indirectly support boys in IGATE-T communities. For context, in Zimbabwe, girls systematically complete less formal education than boys, particularly at the upper secondary and tertiary levels. Despite having achieved gender parity at the primary school level, the gender parity index for lower primary and upper primary school levels falls to 0.95 and 0.87, respectively (Ministry of Primary and Secondary Education, 2020).[5]

This took place through several outreach interventions, including establishing networks of "community champions" who had prominent positions within the communities to advocate for caregivers' involvement in their student's education; holding regular engagement sessions with caregivers and religious leaders to discuss issues around gender and the importance of caregiver (both male and female) support in promoting girls' education; mobilizing religious leaders and other traditional community leaders to raise awareness of traditional gender norms that marginalize girls; and providing training to teachers and headteachers on inclusive, gender-sensitive teaching methods. The teacher training also emphasized the importance of building foundational literacy and numeracy skills early for all students. These interventions are particularly relevant to this context. In rural Zimbabwe, lack of support from community and family members, in addition to gender-specific factors such as early marriage and pregnancy, are associated with higher dropout rates (Surridge et al., 2020; Nordstrom, 2021).[6]

[4]See Nordstrom (2021) for an analysis of the relationships between education outcomes and student characteristics in this data. The analysis uses multiple machine learning methods to describe consistent patterns between student barriers and education outcomes.

[5]This is consistent with broader trends globally, where the gender gap is more visible after students reach adolescence (Leahey and Guo, 2001; Jayachandran, 2015; Klasen, 2018; Gibbs, 2010). These figures represent gender gaps before COVID-19, though the pandemic and associated school closures are expected to have exacerbated existing gender gaps in education (Carvalho and Hares, 2020).

[6]Although I am unable to isolate for attitudes towards these gender-specific factors in the quantitative analysis, I am able to use the qualitative data from focus groups and interviews to examine this.

After COVID-19 led to nationwide school closures, the project also established community learning circles, which were informal education centres led by community champions and IGATE-T trained teachers.[7] Caregiver and community support were the main barriers affecting community learning circle participation, which makes students' participation in the community learning circle an indicator of broader community support.

The first objective of the program was to reshape community attitudes towards marginalized students. A previous study by Cotton et al. (2020) examined the impact of providing information to girls and community members in rural Zimbabwe. Information campaigns are related to attitude interventions since the latter rely on conveying information. However, information alone is generally not sufficient to influence or reshape attitudes. For example, the Breakthrough program studied by Dhar et al. (2021) provided information but also encourages the adolescent participants to reflect on how they could adapt their behaviours based on what they were learning.[8] This is also true of the IGATE-T program studied in this paper.

By addressing fundamental attitudes towards girls and other marginalized students, the theory of change is that the interventions should also ultimately impact outcomes related to education, including progression and learning. Evidence from other contexts indicates that communities that are more supportive of the education of its students have better education outcomes (Burde, 2004; Sebring et al., 2006; Sailors and Samati, 2014). Among other things, these studies show that when communities are more supportive of education, teacher-parent organizations that support students or support school improvements are more functional. This may be particularly true during periods of instability (Burde, 2004). However, the impact of interventions targeting the attitudes of communities that are necessary to bring about this kind of mobilization had not previously been established in the literature.

Building on this, the interventions also aim to promote more inclusive teaching practices within schools and informal learning centres. By encouraging teachers and caregivers to ensure struggling students get the support they need, the theory of change is that this should translate into improved learning and progression outcomes for students

[7]Before the COVID-19 pandemic, the project also established community-based education centres to offer vocational and skills training for out-of-school students. These were targeted to different students and were evaluated using a separate sample. The impact of these community-based education centres for out-of-school youth will not explicitly be evaluated here.

[8]In other literature examining the sources of individuals' attitudes towards gender, Dhar et al. (2019) show that students' attitudes are influenced by the attitudes of their parents, Seguino (2011) show that attitudes are dependent on national religiosity, and Alesina et al. (2013) shows there is a historical basis for modern-day gender norms.

who lack basic foundational reading and numeracy skills that are essential for long-term learning. This is similar to the theory of change underlying Teaching at the Right Level programs, which provide foundational language and mathematics instruction to children based on their ability level, rather than the curriculum assumed based on the student's grade.[9] Skill-based (as opposed to grade-based) approaches like Teaching at the Right Level have been shown to significantly improve students' performance on test scores (Banerjee et al., 2016) and are consistent with the broader evidence around the barriers to good education outcomes for struggling students.[10] In addition to raising awareness of a student's underlying abilities so teachers can use more appropriate teaching methods, this may also alleviate in-class teacher biases.[11] Since the project specifically targets attitudes towards female students and underperforming students, I test for heterogeneous treatment effects along gender and baseline test performance.

After a little over three years of exposure to IGATE-T interventions, there was a 0.403 SD improvement in community attitudes, and this has contributed to participants in the treatment group being 20.9 percentage points more likely to report receiving enough support from their communities to continue learning during the COVID-19 pandemic than students in the control group.[12] I find that communities are not only more supportive of girls' education in general, but that there has been a 0.63 SD improvement in support for adolescent mothers, a subgroup that is significantly more likely to prematurely drop out of school, making attitudes towards adolescent mothers particularly important for supporting their education and quality of life.[13]

[9]These may be particularly relevant in Zimbabwe and other contexts that have automatic progression policies. Such policies advance students to the next grade regardless of their grasp of the material.

[10]In a review of the interventions and policies that affect underperforming adolescents (mostly in North American schools), Cullen et al. (2013) argue that most underperforming students fail because the standard approaches used in schools do not fit the specific needs of struggling students.

[11]For example, there is some evidence that low-performing students receive less attention from teachers (Lipowsky et al., 2007) and that teachers have preferences for teaching higher-performing students in general (Hanushek and Rivkin, 2012).

[12]Throughout the paper I use the term "community members" to refer to the individuals that the project interacted with within IGATE-T communities. This primarily included educators (teachers, headteachers, district education coordinators, instructors at informal training centres), caregivers, parents, religious leaders, community leaders, students, and other citizens residing or working in the IGATE-T communities who interacted with the IGATE-T students such as spouses and in-laws of young girls in the community.

[13]Girls have been legally allowed to re-enter formal schools after having a child since 1999. However, girls who become pregnant still experience significant stigma when they return to school. There is extensive literature looking at the impact of early pregnancy on girls' education and life outcomes, including Buchmann et al. (2017) and Chang et al. (2020), among others.

In addition to improving attitudes and gender norms, the IGATE-T interventions have significantly improved learning and progression outcomes for struggling students, with students who were below the median on literacy tests at the start of the project performing 0.28 SD better than comparison students in the control group by the end of the project. Students who were above the median did not experience any changes in literacy, numeracy, or progression, suggesting that the project's efforts to increase support for marginalized students were effective. This is consistent with findings from the text mining analysis of the qualitative data. The project also led to female students' chore burdens being reduced by 31 minutes per day.

Students who received support to attend the community learning circles during COVID-19 were also 7.8 percentage points less likely to drop out of school and performed better on both literacy and numeracy tests, emphasizing the importance of community and caregiver support during COVID-19 school closures. Together, these findings indicate that community attitudes can be positively influenced by interventions targeting the attitudes communities have towards marginalized students. These interventions also have a positive impact on the education of marginalized students. The precise mechanisms between the impact on attitudes and the impact on education outcomes are less certain; however, the qualitative findings can provide suggestive evidence that the impact on attitudes contributed to the changes in education outcomes.

2. The IGATE-T Program

The IGATE-T project was part of the second phase of the United Kingdom's Foreign, Commonwealth, and Development Office's Girls' Education Challenge. The Girls Education Challenge consists of a collection of 41 projects implemented in 17 countries since 2012. The second phase, referred to as the "transition window," began in 2017.[14] Each project was designed to specifically address the barriers that were relevant to the region being implemented.

To address the barriers that marginalized students in rural Zimbabwe experience, the IGATE-T project engaged with community leaders and caregivers to support students'—particularly girls'—education. Prior to the COVID-19 lockdowns, this focused primarily on formal education options. After the lockdowns began, these engagements encouraged leaders and caregivers to support students' participation in community

[14]More information about the Girls Education Challenge can be found at https://girlseducationchallenge.org/about/.

learning circles while schools were closed. This took place through several simultaneous interventions.

The first involved establishing community champions, who acted as mentors within the community to demonstrate how community members can support student learning. This included working with local leaders and village heads to encourage parents and caregivers during community meetings (discussed below). After COVID-19 closed schools, the network of community champions was mobilized to ensure learning continued during the COVID-19 school closures. Many community champions began facilitating the community learning circles and distributing learning materials to students in their communities who did not have access to learning resources at home.[15] In total, 1,929 people became community champions over the three and a half years the project operated (89% female).

The second intervention involved facilitating community meetings to encourage caregivers to send their children to school. In these meetings, community champions and community leaders (such as the village heads) led discussions to identify issues surrounding education and child protection, particularly for girls and struggling students. In addition to making participants more aware of the barriers students face, community members were encouraged to consider their own attitudes towards girls and marginalized students, and to consider how these influenced their behaviours towards marginalized students. Across the 238 IGATE-T communities, the project set up these meetings in community centres and schools directed at traditional leaders, religious leaders, school development committees, teachers, heads, case care workers (who work for the Zimbabwe Ministry of Social Work and provide support to abuse victims), village health workers, caregivers, school heads, and any other community members that had a role in student education and welfare in the community. These meetings were held quarterly with school development committees in the form of Community "Indabas" (Forums). Additional meetings were held monthly amongst community members and the community champions. Participants were encouraged to share what they learned and discussed with other community members. 29,448 community members participated in the monthly community meetings (65% female) over the three and a half years the project was operating and 3,552 individuals participated in the quarterly Indabas (49% female).

[15]This included daily literacy and numeracy activities, reading cards, study guides, and books that were given to the community learning circle participants, described below.

The third intervention, referred to as the "Whole School Development" model involved providing teacher training to school teachers and headteachers on how to use gender-sensitive and participatory teaching methods and how to help address some of the barriers students experience in classrooms to establish a more inclusive learning environment. The training was delivered every other month by education experts from Zimbabwe and the UK working with World Vision. Like the community meetings, teachers were encouraged to share what they learned from the training sessions with other teachers in their school. In total, 1,717 teachers were trained through the IGATE-T training (52% female) and 319 head teachers were trained (18% female).

The fourth branch of the intervention, which constituted the main part of the project's response to COVID-19 and the subsequent school closures, leveraged the community champions network and the teacher training that had been ongoing between January 2018 (when the project began) and March 2020 (when schools closed due to COVID-19). This took the form of community learning circles (CLCs), which provided community-led informal education options for students in the IGATE-T communities.[16] The CLCs provided an inclusive learning environment to support students who may not have had support for learning at home. The learning circles were run by community champions or teachers who had been trained through the Whole School Development intervention. Teachers and community champions provided support for learning during COVID-19 school closures by offering literacy and numeracy activities, reading cards, and study guides in addition to providing them with tailored instruction to support individual students' learning efforts.[17]

[16]Prior to COVID-19, this kind of community-oriented learning model has been effective in other contexts. For example, Burde and Linden (2013) show that in communities where parents expressed a desire to send both male and female children to school, a similar community-led informal schooling intervention significantly improved both progression and test scores for both boys and girls in rural Afghanistan by removing distance as a barrier to girls who could not travel long distances alone.

[17]When physical meetings could not take place, students or caregivers could pick up the materials from the community champions.

These interventions were designed to improve the attitudes communities had toward marginalized adolescents, including girls and struggling students.[18] According to the theory of change, this set of interventions can be summarized as having two channels of impact. The first is the project's attempt to reshape community attitudes directly through discussion of the barriers that girls and marginalized students experience. Building on this, the second channel involved promoting more inclusive teaching practices within schools and informal learning centres.

By targeting these attitudes, the theory of change predicts that this will make community attitudes less discriminatory against marginalized adolescents, including girls and low-performing students. Although it may seem intuitive that an intervention targeting attitudes would influence attitudes, prior to this research there was no evidence on the efficacy of interventions focused on reshaping community attitudes.

The next assumption in the theory of change, which is supported by evidence in the literature on community support and community mobilization in education (Dhar et al., 2019; Burde, 2004; Sailors and Samati, 2014; Sebring et al., 2006), then predicts that this shift towards more favourable attitudes will lead to behaviours by community members that support education outcomes for marginalized students. This can include reduced chore burdens for marginalized students, increased encouragement for marginalized students (including girls, adolescent mothers, and struggling learners) to attend school and spend time on their studies, and participating in efforts to improve school management. It is well documented that when students have time to attend to their studies and when schools are well managed, students have better education outcomes (Snilstveit et al., 2015). Based on this, the project's theory of change hypothesizes that the change in community attitudes should ultimately improve learning and progression outcomes for these students by affecting behaviours of community members to be more supportive towards the education of these students.

[18]Before the COVID-19 pandemic, the project also established community-based education centres to offer vocational and skills training for out-of-school students. These were targeted to different students and were evaluated using a separate sample. The impact of these community-based education centres for out-of-school youth will not explicitly be evaluated in this paper, however, since out-of-school students have traditionally been marginalized within these communities, it is expected that by normalizing education for out-of-school students, this should also have contributed to shifting attitudes within the communities. This aligns with the project's broader initiative to improve community attitudes toward marginalized girls. Students who were in secondary school at baseline, and are therefore not included in this study, were also given bicycles to reduce their commute times. This intervention has been studied separately in Cotton et al. (2021).

The following sections describe the empirical approach employed to evaluate the impact of the IGATE-T interventions on education outcomes, starting with a description of the data collected for the evaluation. The mixed-methods design allows for conclusions to be made about the *attribution* of the project's impact on attitudes directly, as well as the project's impact on education outcomes, while also examining the mechanisms along this hypothesized theory of change.

3. Study Design and Data

I conducted a quasi-randomized evaluation of the IGATE-T program using a sample of 74 public schools in rural Zimbabwe. To limit the possibility of spillovers to the control areas, this study utilizes a geographic cluster-based randomized design based on the school cluster administrative system used by the Zimbabwe Ministry of Primary and Secondary Education.[19] 238 secondary school clusters were quasi-randomly selected to receive the IGATE-T interventions. Selection was not truly random because schools had to be of sufficient size and had to be public schools to be eligible for the interventions. Four political districts were selected for the evaluation sample, and all treatment schools that had a minimum of thirty students. The set of comparison schools was selected quasi-randomly from the set of schools that were not selected for the original set of the 238 treatment schools in the four evaluation districts. Comparison schools needed to meet the same size criteria as the treatment schools and had to be a sufficient distance from the treatment school clusters to avoid spillover from treatment areas.

Data was collected in three waves by trained teams of local enumerators. Baseline data was collected in October-November 2017, before the IGATE-T interventions began in January 2018. Midline data was collected in May-July 2019, while the interventions were underway. Endline data was collected in May-July 2021.[20] The project stopped implementing interventions in April 2021, though some communities continued to independently implement the community learning circles after the project ended.

3.1. Quantitative Data Sources.

Within each school, survey data was collected from a stratified random selection of girls, with stratification done at the grade level to ensure at least 6 girls were sampled from students in each grade between grades three and

[19]In this context, a "cluster" refers to a geographic area that shares the same secondary school.

[20]Both midline and endline data collection were postponed from the original timeline. Midline data collection was postponed from January 2019 in response to violent protests that led to nationwide school closures that made data collection temporarily unsafe. Endline data collection was postponed from May 2020 in response to COVID-19 travel restrictions and school closures. Both of these disruptions affected treatment and control areas equally, and are not expected to affect the results of this study.

seven.[21] If a girl had a male sibling enrolled at the school, they were also included in the sample to comprise the boy sample. Both male and female students selected were tracked from baseline through to endline, even if they had dropped out of school.

The survey administered to these students collected demographic information, as well as literacy and numeracy assessments. The survey also asked the students about their school's infrastructure, their education aspirations, their leadership capabilities, the support they received from their communities, and other questions about other educational barriers (such as how far they travel to school, and whether they feel safe on the commute). The survey also asked about the student's perceptions of the teaching practices used in their classrooms.

The literacy and numeracy assessments included the Early Grade Reading Assessment (EGRA) and Secondary Grade Reading Assessment (SeGRA) to evaluate the student's literacy skills. The test also included the Early Grade Mathematics Assessment (EGMA) and the Secondary Grade Mathematics Assessment (SeGMA) to evaluate the student's mathematics skills.[22] Due to restrictions around COVID-19, learning assessment data was only collected from students in Grade 7 and Form 4 at endline.[23] This means that the sample size available at endline to evaluate learning outcomes is smaller than the sample available at midline.

In addition to collecting survey data from the students themselves, data was also collected from girls' primary caregivers to get additional information about the girls' home environment. Each girl's teacher was also surveyed to get additional information about the girl's in-class environment. Surveys with the caregivers and teachers were not conducted at endline due to the project's budget constraints after COVID-19. Caregiver and teacher surveys were not collected for boys at any time period.

[21] At baseline, students in Form 1 and Form 2 (equivalent to grade eight and nine, respectively) were also selected. However, after the delay to the midline data collection, it was decided these students would be dropped from the sample after midline since they would be too old to be sampled by endline. Given the significance of the midline to endline period—which includes the COVID-19 pandemic and the project's response to it—these cohorts have been excluded for the purposes of this study to focus on students who are included in the endline data collection.

[22] The EGRA/SeGRA and EGMA/SeGMA assessments are widely used assessments developed for the United States Agency for International Development to assess literacy and numeracy skills. The assessments include multiple "subtasks" that increase in difficulty to assess more advanced students. Students were given the subtasks that were relevant to their grade level, plus the subtasks that were three grades above their baseline grade to allow for changes to be measured as students progress through their education during the life of the project. The assessments were adjusted and piloted to ensure they were culturally appropriate using the standard EGRA/EGMA guidelines. Assessments were also calibrated at each midline and endline to ensure they were equally difficult across time.

[23] Data collection teams followed all local health guidelines during endline data collection.

Headteacher surveys were collected at all three time periods from every school in the control and treatment areas. These surveys were used to collect data on the child protection measures in place at the school, and to identify other support schools receive to be aware of any possible contamination sources. Additional support from NGOs or community organizations was rare, and there was no systematic relationship between the types of support schools received and the school's treatment status.

Table 1 displays summary statistics for treatment and control samples at baseline. Differences in means are also reported. There are no significant differences in the demographic composition of the samples, nor are there any significant differences between the baseline test scores or attendance. The only significant difference is in students' daily chore burden, which is slightly higher in the treatment group at baseline. It is unclear what explains this baseline difference. However, it has been controlled for when ordinary least squares specifications have been used. This is described in more detail in section 4.

Table 1. Summary Statistics

	Control	**Treatment**	**Difference**
Grade	4.831	4.765	-0.066
Age	10.741	10.865	0.124
Female	0.859	0.867	0.008
Pregnant or parent	0.002	0.002	0.000
Disability	0.094	0.096	0.002
Lives without parent	0.295	0.266	-0.030
Orphan	0.153	0.151	-0.002
Daily chores (hrs)	1.476	1.828	0.352*
HH Experiences Hunger	0.369	0.374	0.005
PCG has no education	0.095	0.103	0.008
Apostolic	0.347	0.391	0.045
Safe Commute	0.771	0.794	0.023
Teacher frequently absent	0.264	0.278	0.014
No water at school	0.245	0.232	-0.013
Literacy Test Score	37.006	36.171	-0.834
Numeracy Test Score	57.471	56.409	-1.063
YLI Score	54.472	54.721	0.249
Attendance (/22 days)	15.224	14.592	-0.632
Number of clusters	37	37	0
Observations	611	617	6
Attrition Rate (%)	43.378	39.091	-4.287

Note: * $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$

1,719 students in grades 3 through 7 were interviewed at baseline from the 74 schools included in the evaluation sample (37 treatment schools and 37 comparison schools). Interviews were conducted at the school, or at their homes if they could not be contacted at the school. At endline, 1,228 of these students could be recontacted. As shown in Table 1, this amounts to an attrition rate of approximately 40% for both the control and treatment groups over the three and a half year period. Importantly, attrition rates did not significantly differ between the treatment and comparison areas. This is discussed in more detail in Appendix B, where in addition to showing the similarities in baseline characteristics between the attrited and recontacted subgroups, I apply regularization methods—including lasso and elastic net—to all observable characteristics in the data to identify the features that are the most important predictors of attrition.

Notably, treatment status is not identified as a significant predictor of attrition using either lasso or elastic net.[24] Using these regularization techniques allows for a comprehensive study of the relationships between attrition and the over 1,500 observable characteristics available for each girl at baseline. The results suggest that apart from age—which is a well-established predictor of school dropouts—there are very few characteristics associated with attrition rates. Furthermore, none of these characteristics are associated with treatment status, which suggests that attrition should not bias the findings of this study.

3.2. **Qualitative Data Sources.** At all three time periods, qualitative interviews were conducted with a random subset of the girls, caregivers, and teachers surveyed. Additional qualitative interviews were conducted with community champions and with specific officials, such as the District School Inspectors, and religious leaders in the communities. These interviews took the form of either key informant interviews or focus groups and were designed to understand the barriers and enablers of education in IGATE-T communities. Unlike the quantitative data, the qualitative data is not necessarily collected from the same people at baseline, midline, and endline. A diverse set of perspectives was included in the qualitative sample to represent different stakeholders

[24]In the lasso specification, which applies a stricter penalty term (selected using one-dimensional cross-validation), I find that there are no characteristics—out of more than 1,500 possible baseline features in the data—that improve the ability to predict attrition rates in the test sample beyond the average attrition rate. When the penalty for complexity is relaxed using elastic net (where the penalty terms were selected using two-dimensional cross-validation) I find that being in a higher grade at baseline, as well as living in one of the four evaluation districts or having a head of household with no education were identified as significant predictors of attrition. However, including these characteristics to explain attrition rates did not improve the explanatory power over the model including just the intercept. See Table 14 in Appendix B and the accompanying discussion for more detail on this analysis.

within the IGATE-T communities. Unless otherwise specified, the qualitative analysis will consider the perspectives of all these community members together to reflect community attitudes. The total number of interviews conducted at each time period is presented in Table 2. The qualitative data is analyzed at a question level. However, since the data comes from verbatim transcriptions of the interviews and focus groups and the questions may not be worded exactly the same way by each enumerator, I rely on fuzzy matching to identify the similar questions asked across different interview types at each sample point.

Table 2. Qualitative Sample Size

	Baseline	**Midline**	**Endline**
IGATE-T Evaluation	71	90	69

Qualitative data was only collected from treatment group areas, so I do not have counterfactual evidence on the changes in community attitudes when I am evaluating the changes in community sentiment measured within these interviews. As I will discuss in section 4.1, the evidence from the text mining analysis of this qualitative data is used primarily to gain insights into the specific community attitudes that have changed since the start of the project; and to identify patterns in the types of terms that are used in responses to open-ended questions that can identify relevant channels or barriers. This is supplemented by the findings of the quasi-randomized field experiment to get a more comprehensive understanding of the mechanisms by which IGATE-T has had an impact on attitudes and education outcomes.

4. Empirical Methodology

The intervention is hypothesized to make community attitudes less discriminatory against marginalized adolescents, including girls and low-performing students. The theory of change posits that positive changes in community attitudes should improve community attitudes, and that this should ultimately improve learning and progression outcomes for these students by affecting behaviours of community members to be more supportive towards the education of these students. In this section, I describe the estimation strategy used to test this hypothesis before presenting the results. The identification strategy employed here allows me to measure the impact of the IGATE-T interventions on community attitudes, and the impact on education outcomes (separately), while using mixed-methods approaches to examine possible mechanisms.

4.1. **Text Mining Methods.** I employ text mining methods to analyze the qualitative interviews in two ways. The first involves examining changes in community attitudes since the IGATE-T interventions began. Sentiment analysis, which involves measuring the "emotional intent" of text data, provides a useful way to measure attitudes toward sensitive topics (Wilson, 2008; Xu et al., 2019; Liu et al., 2010), and provides a measure of implicit attitudes. These tools offer an effective measure of the opinions or attitudes of individuals (Karlgren et al., 2012), which make them particularly relevant for research on attitude interventions.

The second approach used in this paper relies on an analysis of the specific terms used in response to open-ended questions. The open-ended nature of qualitative interviews makes it possible to develop a more comprehensive understanding of the mechanisms through which the IGATE-T interventions affect learning and progression.[25] Since the qualitative data is only collected from the treatment areas, these insights will not provide causal estimates of the project's impact. However, when complemented by the findings from the impact analysis conducted on the survey data, adopting this kind of mixed-methods approach is an effective way to explain the mechanisms in the theory of change (White, 2013; Creswell and Poth, 2016).

4.1.1. *Sentiment Analysis.* Sentiment analysis is a useful tool for measuring the attitudes and opinions expressed in text data. While often used to analyze large volumes of text data from social media posts (Agarwal et al., 2011; Rui et al., 2013), academic writing (Lennox et al., 2020; Solovyev et al., 2019), news articles (Yu et al., 2013; Balahur et al., 2013), literature (Mohammad, 2012; Jacobs et al., 2020), or customer reviews (Xianghua et al., 2013; Hu and Li, 2011; Rui et al., 2013), the same methods can be applied to text data from transcriptions of qualitative interviews. These tools are particularly useful for measuring attitudes regarding sensitive topics since they can detect the context around the content in qualitative data (Xu et al., 2019; Liu et al., 2010; Neviarouskaya et al., 2010). The relatively large volume of qualitative data described in section 3.2 for this project makes it possible to employ these methods in this study.

This paper adopts a dictionary-based approach to sentiment analysis, which uses "lexicons" to find the sentiment associated with individual words in interview responses

[25]This use of qualitative data is a text mining approach to qualitative content analysis, which is a widely-used qualitative research technique to interpret meaning from text data that relies on manually reviewing and coding themes in the data (Mayring, 2004).

(Silge and Robinson, 2017). These lexicons assign an emotional intent to words used in the English language. I combine four common lexicons for this analysis, including the NRC lexicon (Mohammad and Turney, 2013), the AFINN lexicon (Nielsen, 2011), the Bing lexicon (Hu and Liu, 2004), and the syuzhet lexicon (Jockers, 2017), which each use a different approach to assign an emotive value to English words.[26,27] These have been integrated by Rinker (2017), who combined the four different emotion assignment approaches into one scale—measured between -1 and +1—and incorporates weighting for negation terms.

Negation terms (also referred to as "valence shifters") are an important part of speech that change or amplify the meaning of terms. For example, "not good" would negate the positive emotions associated with "good", while "very good" would amplify these emotions. To account for these, I use the augmented dictionary approach introduced by Rinker (2017), which adds weighting for valence shifters to amplify, deamplify, or reverse the impact of terms with emotional intent in a phrase. All of the qualitative interviews analyzed in the data contain at least one type of valence shifter, making this weighting particularly relevant.[28] Next, I describe the algorithm to measure sentiment using this approach.

4.1.2. *Sentiment Analysis Algorithm.* Using the approach introduced by Rinker (2017), I first identify all emotive words within a sentence in the response. The four words preceding each emotive word and two words following the emotive word are then analyzed to look for valence shifters within this seven-word "cluster".

The emotional polarity of the cluster is calculated by increasing the value of the emotive word when there are amplifiers present, decreasing if there are de-amplifiers present, and negating if there is a negation term present. Adversative conjunctions such as "but" and "however" also adjust the polarity of the original emotive word if

[26]NRC uses binary indicators for whether a word fits into different emotive categories such as positive, negative, sadness, joy, trust; the Bing lexicon categorizes words as positive or negative; the AFINN lexicon assigns words a value between -5 and +5, with negative (positive) scores representing a negative (positive) sentiment; the syuzhet lexicon uses a scale of -1 to +1, with sixteen values in between where negative (positive) scores represent negative (positive) sentiment.

[27]Most of the interviews were conducted in English, with some exceptions for caregivers and young students. These interviews were translated verbatim into English. English is one of sixteen official languages spoken in Zimbabwe, and approximately 90% of the population is fluent in English, which is the language of instruction in Zimbabwean schools.

[28]This is common in other applications of sentiment analysis (Yu et al., 2016; Kennedy and Inkpen, 2006), with negators appearing around approximately 20% of emotive words, and amplifiers appearing around approximately 10% of emotive words (Naldi, 2019).

they are observed within the cluster. Amplifiers and deamplifiers increase or decrease the polarity by a factor of 0.8, while negation terms will reverse the sign of the emotive word, and adversative conjunctions up-weight the cluster by a factor of 0.85 to account for the fact that adversative conjunctions assign greater value to the next clause in a phrase.[29]

The overall polarity score of each sentence is then calculated as the sum of all weighted clusters, divided by the square root of the total word count. The average sentiment within an overall response is then calculated as the mean sentiment score across all words.[30] As one of the only approaches to sentiment analysis that properly adjust sentiment to account for valence shifters, this algorithm has been widely used to study teacher sentiment towards student behaviour (Chen et al., 2020), attitudes of UK energy consumers (Ikoro et al., 2018), trends in attitudes of researchers publishing on conservation science (Lennox et al., 2020), and biases in media coverage (Dalal et al., 2019).[31]

4.2. **Impact Evaluation using Survey Data.** There are two types of outcomes that will be evaluated from the survey data. The first is continuous outcomes, which include test scores and chore time. The second is binary outcomes, which include whether a student has dropped out or repeated a grade by midline or endline, and whether a student received support from their family or their community to continue learning during pandemic-related school closures.[32] To estimate the impact of IGATE-T on continuous outcomes, I use a difference-in-differences framework using the following two-way fixed-effects specification:

$$y_{it} = \beta_0 + \beta_1 treatment_i \times after_t + \mathbf{X_{it}\Gamma} + \alpha_i + \alpha_t + \epsilon_{it} \tag{1}$$

[29] I use the default weightings recommended by Rinker (2017) for each of these shifter types.

[30] For additional details on this algorithm, see Rinker (2017).

[31] One possible limitation of this approach is that the phrase-level methodology may overlook negation that is not local (Wilson et al., 2005). For example "not happy" represents local negation while "does not seem very happy" is an example of longer-distance negation within a phrase. The latter would not be recognized as a negation using this approach if the term is more than four words away from the emotive word analyzed since the sentiment analysis algorithm used here looks at seven-word clusters. Since the algorithm also calculates overall sentiment analyzed within an entire question response, this should account for negation across phrases. Moreover, since this would be an issue for both baseline and endline measures, this should not affect the interpretation of the results in this application.

[32] As shown in Nordstrom and Cotton (2021), analyzing both progression and learning outcomes is essential for understanding the full impact of education interventions in this environment.

where y_{it} represents student i's literacy or numeracy test scores[33] or chore time (hours), $treatment_i$ indicates whether the student comes from a treatment school, and $after_t$ is an indicator for whether the data refers to an observation collected before or after IGATE-T interventions began. The results in section 5 include baseline to midline and baseline to endline results to show the impacts of the project over time. Additional results are also shown between midline to endline to examine how the impacts are distributed. In these specifications, $after_t$ is equal to 0 for the midline observations and 1 for the endline observations.

The coefficient of interest, β_1, measures the impact of the IGATE-T interventions on literacy/numeracy or chore times. Assignment to the IGATE-T intervention group was quasi-random with minimal selection criteria, the estimation can be interpreted as *causal*. However, since the interventions were not taken up by all students in treatment locations, β_1 should be interpreted as a causal estimate of IGATE-T's intent-to-treat (ITT). This is particularly well suited to the program design, which encouraged information sharing within the treatment schools and communities.

X_{it} represents a matrix of student-specific variables that are allowed to vary over time, including student i's age, grade, and chore burden at time t, respectively. Additionally, results from additional ordinary least squares (OLS) regressions without controls or fixed effects are included for comparison. In the OLS specification with controls, control variables for baseline characteristics including the student's district, household socioeconomic status, chore burden, and caregiver education level have also been included.

To estimate the effects of the IGATE-T program on discrete education outcomes, I estimate the following non-linear specification:

$$Pr(y_i = 1|\boldsymbol{X_i}) = F(\alpha_1 treatment_i + \boldsymbol{X_i \Omega}) + \epsilon_i \tag{2}$$

where y_i is an indicator for whether student i has dropped out of school or repeated a grade by or midline or endline, or received support to continue learning during the COVID-19 school closures by endline. Like the continuous outcomes, the project's impact at both midline and endline is presented to show the timing of impacts. X_i includes a set of baseline characteristics, including the student's age, grade, district, household socioeconomic status, and caregiver education level. Again, the coefficient of interest,

[33] The literature around gender gaps in education find that gaps in literacy and numeracy vary as girls get older, and depend on their school and community's expectations for girls' future opportunities (Marks, 2008). For this reason, both literacy and numeracy changes are studied in this paper.

α_1, can be interpreted as a causal estimate of the project's ITT. Since randomization to the treatment and control groups took place at the school level, standard errors have been clustered at the school level for both continuous and discrete specifications.

Since the project specifically targets attitudes towards female students and underperforming students, I test for heterogeneous treatment effects along gender and baseline test performance, which is a proxy for baseline academic ability. The literacy and numeracy test scores have been standardized and adjusted to a student's grade level. Heterogeneous effects are not included for all analyses due to limited sample sizes for some measures. The is particularly true for gender disaggregation. The sample consists of a little over 10% boys, making it significantly smaller than the girl sample (see table 1). At endline, literacy and numeracy tests were only collected from a subset of the sample. This is not a concern for the overall sample, which is still sufficiently powered, but this significantly limits the sample's ability to detect changes in boys' learning outcomes. Gender disaggregation is also difficult for the progression outcomes, due to the small number of boys overall combined with the small number of students who drop out of school in this age group.

To examine the relevant mechanisms, I consider the differences between those who received enough support from their communities to continue learning during school closures and those who did not. This indicates how community support may have contributed to behaviour changes observable by students, and the results have been disaggregated along this indicator to understand how community support has affected education outcomes. The second approach I use to examine the channels through which community support affects education outcomes is by examining the differences in outcomes by community learning circle participation. Caregiver and community support were the main barriers affecting community learning circle participation rates, which makes students' participation in the community learning circles an additional indicator of broader community support. Given that the set of outcomes and heterogeneity analysis selected for this analysis is parsimonious and highly relevant to the theory of change, I do not make adjustments for multiple hypothesis testing.

5. Results

This section presents the effects of the IGATE-T program on the attitudes of communities and the education outcomes of its participants, starting with an analysis of the changes observed in community attitudes.

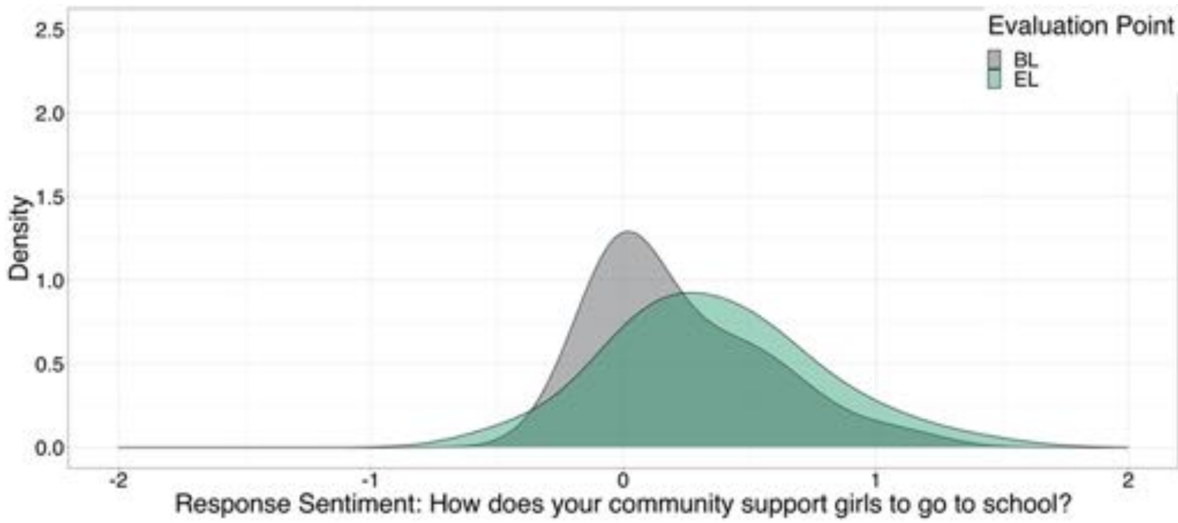
5.1. **Impact on Community Attitudes.** The IGATE-T project encouraged community leaders and caregivers to be more supportive of students'—particularly girls'—efforts to study and attend school, and then to attend community learning circles when schools were closed due to COVID-19. Here I examine how community attitudes have changed since the beginning of the IGATE-T project. Community attitudes towards girls' education play an important role in determining girls' education outcomes. Caregivers make decisions about how to allocate the household's economic resources, which determine whether or not girls can attend school since the fees and levies associated with attending school in Zimbabwe make up a significant portion of household earnings, particularly in rural areas. Religious communities also play an important role in establishing the community's norms around early marriage and supporting adolescent mothers to return to school. Girls who become pregnant, or who have been victims of abuse or gender-based violence, experience significant stigma from community members and student peers, which make many students reluctant to return to school.

IGATE-T specifically works to address these barriers by encouraging caregivers, community leaders, and religious communities to support adolescent mothers and not to condone marriages that involve girls who have not completed school before they are of legal age.[34] The project also helped schools establish child protection committees, which gave parents, religious leaders, and educators the information and resources needed to respond to reported cases of abuse or violence. If these efforts have been successful, we should expect more positive community attitudes towards girls' education, adolescent mothers, and victims of abuse since baseline. I find evidence of economically and statistically significant improvements in community attitudes since baseline, as shown in the following tables and figures.

Figure 1 shows the changes in community sentiment when asked about how community members support girls to go to school. At the end of the project, there was a large, positive shift in the attitudes held by community members when they were discussing how their communities support girls' education. Compared to sentiment levels measured at baseline, communities are 58% more likely to have positive attitudes towards their community's support for girls' education, representing a 0.403 SD improvement in community attitudes.

[34]For context, the legal age of consent in Zimbabwe was 16 when the interventions began in 2018. In 2019 the country was in the process of reviewing the age of consent and considering legislation to increase it to 18 (Musa Kika, 2019). However, enforcement of the legal age is inconsistent and is at odds with traditional practices observed in some communities.

Figure 1. Reported Sentiment Associated with Community's Support for Girls' Education



Although there was an average positive response to begin with at baseline, a significant mass of individuals had very neutral attitudes towards the community's role in supporting girls' education. Since baseline, this is reflected in a shift of the overall distribution towards more positive attitudes. Specifically, the distribution of attitudes has almost entirely shifted right,[35] with a greater mass in the long tail of positive attitudes. This has contributed to an economically and statistically significant improvement in the attitudes communities hold about supporting girls' education since baseline.[36] By relying on the *sentiment* associated with the language interviewees use, rather than just looking at the content of these interviews, I avoid biases that may arise if individuals have expectations about what the intentions of the enumerators are, which may lead to more favourable explicit attitudes being reported.[37] Instead, these measures allow for a measure of individuals' implicit gender attitudes.

This shift in community attitudes to girls' education is reflected in significant changes that are observable to students. Table 3 shows that, at endline, participants in the treatment group were 20.9 percentage points more likely to report having received enough support from their communities and 7.8 percentage points more likely to report having received enough support from their families to continue learning during the

[35]The p-value on the Kolmogorov-Smirnov test of the null that the baseline and endline sentiment distributions are equivalent is 0.002.

[36]The p-value on a t-test of the null that the two distribution means in Figure 1 are equivalent is 0.02.

[37]Although all enumerators were transparent about their independence from the IGATE-T project, it is plausible that this distinction would not have been obvious to interviewees.

COVID-19 pandemic and associated year-long school closures, compared to students in the comparison group. In section 5.3 I show weak but suggestive evidence that this may have contributed to a decline in dropout rates, but not test scores.

TABLE 3. Impact of IGATE-T on Reported Support from Community and Family Through the COVID-19 Pandemic

	Received enough support to continue learning during pandemic from:	
	Family	Community
Treatment	0.078**	0.209***
	(0.038)	(0.033)
Controls	✓	✓
Students	614	614

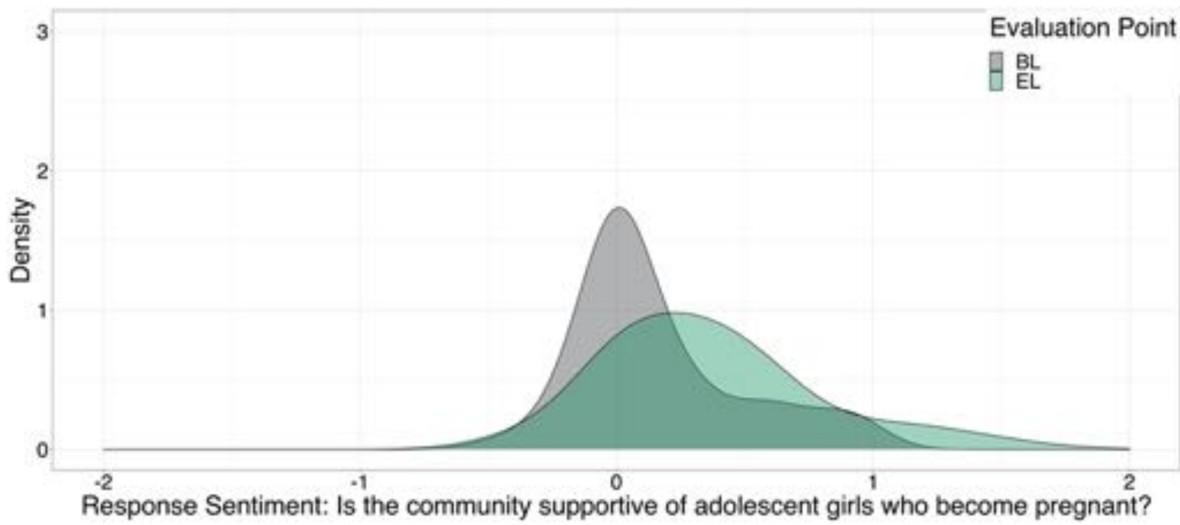
Note: Cluster-robust standard errors are in parentheses. Robust standard errors are clustered at the school level. Controls in baseline characteristics of the learner's age, grade, chore burden, district, household socioeconomic status (measured by the household's ability to pay for basic needs), and caregiver education level. * $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$

While these student reports describe the impact that IGATE-T has had on community attitudes compared to a counterfactual, the sentiment analysis allows for a more precise understanding of how community attitudes have changed (around young mothers and abuse victims, for example). Qualitative interviewees were also asked about how communities support adolescent mothers and girls who become pregnant. The respondents' sentiment to this question at baseline and endline is shown in Figure 2.

After the project, I find that community members also hold significantly more positive attitudes towards adolescent girls who become pregnant. This improvement is both economically and statistically significant, representing an 0.63 SD improvement in attitudes towards adolescent girls who become pregnant, overall.[38] This is particularly important as pregnancy is one of the leading predictors of dropping out of school in Zimbabwe (Surridge et al., 2020; Nordstrom, 2021).

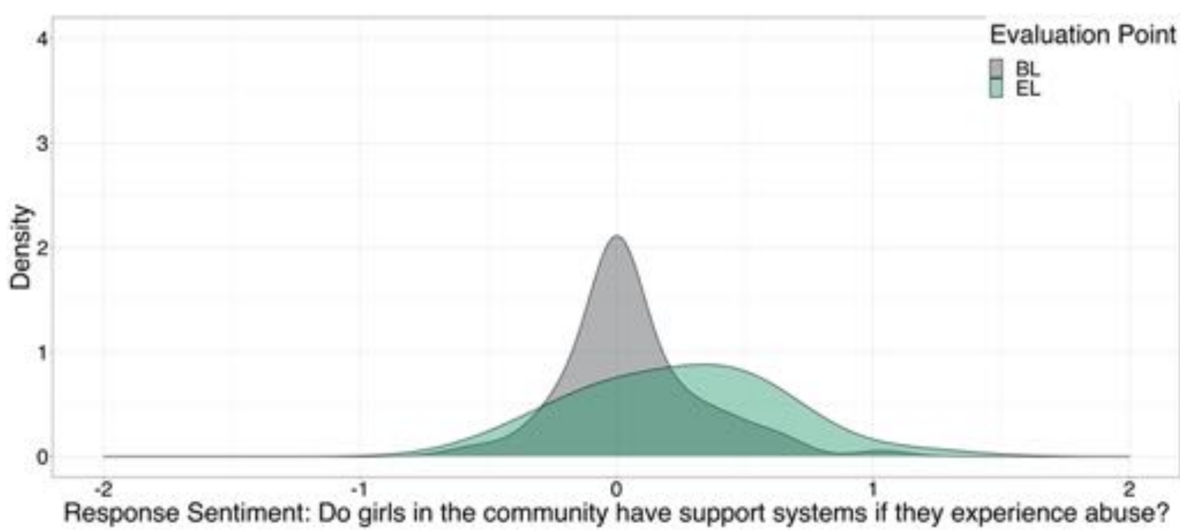
[38] The p-value on both a Kolmogorov-Smirnov test and a t-test of the distribution means are both < 0.001.

Figure 2. Reported Sentiment Associated with Community's Support for Adolescent Mothers



To examine the changes in community attitudes towards girls who have been victims of abuse—who are traditionally victims of significant stigmatization within communities—Figure 3 shows that, like attitudes towards overall girls' education and to adolescent mothers, attitudes towards girls who have been victims of abuse or violence have become more positive within the community.

Figure 3. Reported Sentiment Associated with Community's Support for Victims of Abuse



Together, these results suggest that communities have become more supportive of girls' education, particularly towards girls that have traditionally been among the most marginalized in these communities. The qualitative data is not collected from control areas, so we cannot explicitly make conclusions about the causality of this change in community attitudes from the sentiment analysis alone. However, these descriptive changes are consistent with the increased support students reported receiving during COVID-19 school closures from their communities and their families, compared to the control group. We can also consider these attitude shifts to measurable behaviour changes in the quantitative data to confirm these trends, as discussed in the following subsection.[39]

5.1.1. *Chore Allocations.* One indicator of community support for children's education is how households allocate chore burdens. Before IGATE-T began, traditional gender norms typically contributed to girls having higher chore burdens than boys in IGATE-T communities, limiting the time they had available to attend to their studies. At baseline, learners were doing approximately one a half hours of daily chores, on average. High chore burdens primarily affect learners' education outcomes by limiting their time available for homework or classes. The IGATE-T project specifically worked with communities and schools to address this by encouraging caregivers to notice how chores were affecting students—particularly girls—and to allocate chores more equitably within a household. Using the specification described in equation 1, Table 4 shows that IGATE-T led to students doing 0.4 fewer hours of chores each day by endline, equivalent to 24 minutes each day.

Table 5 shows that this impact is isolated to girls who were part of the IGATE-T program, who are doing 31 fewer minutes of chores each day after the project. Although we do observe an increase in chore burdens experienced by boys, which would be expected if the chores are being reallocated from female to male children within the same households, this increase for boys is not statistically significant.

This is consistent with the broader change in community attitudes that have been shown above and may suggest that changing community attitudes have facilitated tangible changes in behaviour that promote improved educational outcomes for girls in these communities. Regardless of the relationship between attitude shifts and the change in chore burdens, the findings presented so far indicate that the IGATE-T program has

[39]One such measure of these changing attitudes is the frequency at which abuse cases are reported in treatment versus control schools. At endline, 25% of headteachers in treatment schools stated that their school's child protection committee had been notified of a case of abuse, compared to 13% of headteachers in control schools. Rather than being a signal of increasing cases of abuse in treatment areas, it is more likely that this is an indication that community members are more *aware* of abuse when it takes place and are more willing to provide support to the victims by reporting it. This is confirmed by the qualitative content analysis of the interview data.

Table 4. Impact of IGATE-T on Chore Burdens

	Chore Time (hours)		
	(1)	(2)	(3)
EL - BL DiD			
Treatment x Time	-0.394**	-0.415*	-0.414
	(0.178)	(0.174)	(0.259)
Controls		✓	✓
Time and Learner FE			✓
ML - BL DiD			
Treatment x Time	-0.325	-0.352	-0.382
	(0.249)	(0.243)	(0.351)
Controls		✓	✓
Time and Learner FE			✓
EL - ML DiD			
Treatment x Time	-0.069	-0.070	-0.036
	(0.216)	(0.213)	(0.313)
Controls		✓	✓
Time and Learner FE			✓
Students	1,228	1,228	1,228

Note: Cluster-robust standard errors are in parentheses. Standard errors are clustered at the school level. Controls in OLS specifications (panels 1 and 2) include the learner's age and grade as well as baseline characteristics of the learner's district, household socioeconomic status (measured by the household's ability to pay for basic needs), and caregiver education level. * $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$

Table 5. Impact of IGATE-T on Chore Burdens by Gender

	Chore Time (hours) - Girls			Chore Time (hours) - Boys		
	(1)	(2)	(3)	(4)	(5)	(6)
EL - BL DiD						
Treatment	-0.464**	-0.512**	-0.512*	-0.0014	0.161	0.109
	(0.207)	(0.202)	(0.301)	(0.331)	(0.317)	(0.462)
Controls		✓	✓		✓	✓
Time and Learner FE			✓			✓
Students	1,066	1,066	1,066	162	162	162

Note: Cluster-robust standard errors are in parentheses. Standard errors are clustered at the school level. Controls in OLS specifications (panels 1, 2, 4, and 5) include the learner's age and grade as well as baseline characteristics of the learner's district, household socioeconomic status (measured by the household's ability to pay for basic needs), and caregiver education level. * $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$

contributed to meaningful, positive changes in community attitudes as well as changes in behaviours within households that can support girls' education. In the following sections, I discuss how IGATE-T has affected students' overall learning and progression.

5.2. **Impacts on Test Scores and Student Learning.** Columns 3 and 6 in Table 6 shows the coefficient estimates from the identification strategy presented in equation 1.[40] I find no evidence that overall test scores have changed for learners in IGATE-T treatment schools between baseline and endline or between baseline and midline.

Table 6. Impact of IGATE-T on Test Scores

	Literacy (SD)			**Numeracy (SD)**		
	(1)	(2)	(3)	(4)	(5)	(6)
EL - BL DiD						
Treatment x Time	0.193	0.106	0.104	-0.056	0.070	0.0067
	(0.136)	(0.170)	(0.130)	(0.146)	(0.165)	(0.216)
Controls		✓	✓		✓	✓
Time and Learner FE			✓			✓
Learners	266	266	266	266	266	266
ML - BL DiD						
Treatment x Time	0.0046	0.104	0.063	-1.021	-0.945	-0.456
	(1.191)	(0.976)	(1.826)	(0.146)	(1.130)	(0.216)
Controls		✓	✓		✓	✓
Time and Learner FE			✓			✓
Learners	1,228	1,228	1,228	1,228	1,228	1,228
EL - ML DiD						
Treatment x Time	0.202**	0.234*	0.177	0.146	0.235	0.160
	(0.103)	(0.135)	(0.145)	(0.154)	(0.178)	(0.221)
Controls		✓	✓		✓	✓
Time and Learner FE			✓			✓
Students	266	266	266	266	266	266

Note: Cluster-robust standard errors are in parentheses. Standard errors are clustered at the school level. Controls in OLS specifications (panels 1, 2, 4, and 5) include the learner's age and grade as well as baseline characteristics of the learner's district, household socioeconomic status (measured by the household's ability to pay for basic needs), and caregiver education level. * $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$

Between midline and endline, I observe a positive (0.188 SD) impact on test scores between midline and endline for literacy tests. This may be an indication that the project's response to COVID-19—which took place between midline and endline—was particularly effective for improving literacy outcomes. This may also be due to the interventions that were implemented post-midline or were delayed responses to the

[40]Columns 1, 2, 4, and 5 present the results of a standard OLS estimation with and without controls, as indicated in section 4.

interventions that took place before midline. The midline to endline results are not statistically significant when we employ the fixed-effects identification strategy presented in equation 1. The effect size remains relatively similar to the OLS specification with controls, which suggests this may be a consequence of the relatively small sample size at endline for test scores.

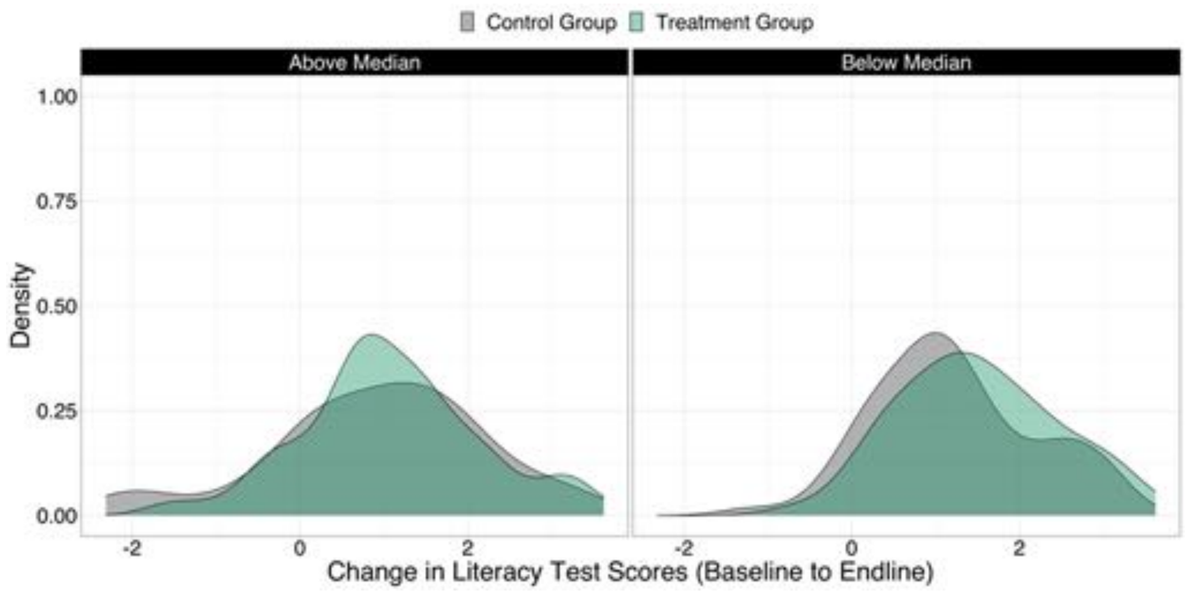
To understand the impact that IGATE-T has had on test scores, I measure several heterogeneous effects. The first is the project's impact on supporting struggling students. The IGATE-T interventions were specifically designed to target students who were struggling with basic literacy and numeracy since these students are often marginalized within the education system and have poorer education outcomes (Surridge et al., 2020). Specifically, the teacher training provided by IGATE-T equipped teachers with techniques to identify and support students who were struggling with basic skills. The project also encouraged community leaders and community members to support students who may be struggling academically, instead of encouraging them to either drop out of school or begin working before completing their education.

Figure 4 shows the change in literacy test scores between baseline and endline for those who were above or below the median at baseline. The left panel in Figure 4 shows that the project did not have a significant impact on participants who were already performing above the median at baseline (p-value on a t-test of a null that there is no difference in means = 0.4). The project did have a significant, positive impact on participants who were below the median at baseline, compared to the control group (p-value on a t-test of the null that there is no difference in means= 0.0992). This is shown in the right-most panel of Figure 4.

This is evidenced by the results in Table 7, which adds an interaction term between the IGATE-T treatment-time indicator and the baseline literacy or numeracy test scores to the specification in Equation 1 to estimate the heterogeneous effects of the IGATE-T interventions across students with varying ability at baseline.

The coefficient on the interaction term is negative, which indicates that the efficacy of the intervention is inversely related to the baseline test scores of the participants. Specifically, for every one percentage point worse a student did at baseline, IGATE-T would be expected to have improved numeracy scores by an additional 0.03 SD at endline. The effect on literacy is also inversely related to baseline literacy ability, however, it is not statistically significant after fixed effects are added. This may be due to limited explanatory power, since we see that the effect is still significant for numeracy

FIGURE 4. Impact on Changes in Literacy Test Scores by Baseline Proficiency



between baseline and midline when a larger sample was available. This means that the impact of the IGATE-T program is *strongest* among those who were doing *worst* at the start of the program, which is consistent with the project's theory of change.

These results also speak to changes that specifically took place for girls. The boy learning assessments sample is not large enough to be analyzed separately. However, as is expected given the small size of the boy test score sample, the results are largely the same when only girls are included in the specification used for table 7. Together with the rest of the findings presented so far, this indicates that the IGATE-T interventions were effective for marginalized students in facilitating improvements in both literacy and numeracy. In the following subsection, I also show that IGATE-T is also most effective in improving progression rates for the students who performed weakest at baseline.

If the theory of change is true, we would expect that students improved most on the most basic skills. Figures 5 and 6 show the parameter estimates from the heterogeneity analysis for each subtask in the literacy and numeracy tests, respectively. The plots show the coefficient of the interaction between the treatment and baseline test scores, with 95% confidence intervals and provide evidence that the project's impact on the weakest students is concentrated among the simplest tasks for both literacy and numeracy, particularly between baseline and midline, when compared to comparability weak students in the comparison group.

Table 7. Heterogeneous Impact of IGATE-T on Test Scores By Baseline Test Performance

	Literacy (SD)		**Numeracy (SD)**	
	(1)	(2)	(3)	(4)
EL - BL DiD				
Treatment x Time	0.550*	0.444	1.625***	1.565***
	(0.279)	(0.293)	(0.386)	(0.340)
Treatment x Time x Baseline Test Score	-0.00947*	-0.00882	-0.0290***	-0.0269***
	(0.00537)	(0.00577)	(0.00511)	(0.00467)
Controls		✓		✓
Time and Learner FE	✓	✓	✓	✓
Learners	266	266	266	266
ML - BL DiD				
Treatment x Time	0.321*	0.301	1.159***	1.224***
	(0.171)	(0.188)	(0.317)	(0.361)
Treatment x Time x Baseline Test Score	-0.00874**	-0.00908**	-0.0235***	-0.0226***
	(0.00356)	(0.00450)	(0.00472)	(0.00539)
Controls		✓		✓
Time and Learner FE	✓	✓	✓	✓
Learners	1,228	1,228	1,228	1,228
EL - ML DiD				
Treatment x Time	0.230	0.197	0.466	0.492
	(0.203)	(0.208)	(0.433)	(0.451)
Treatment x Time x Baseline Test Score	-0.000731	-0.000541	-0.00551	-0.00570
	(0.00379)	(0.00393)	(0.00554)	(0.00574)
Controls		✓		✓
Time and Learner FE	✓	✓	✓	✓
Students	266	266	266	266

Note: Cluster-robust standard errors are in parentheses. Standard errors are clustered at the school level. Controls in OLS specifications (panels 1, 2, 4, and 5) include the learner's age and grade as well as baseline characteristics of the learner's district, household socioeconomic status (measured by the household's ability to pay for basic needs), and caregiver education level. * $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$

The impact of the IGATE-T interventions is not observable for any subtask individually when we look at the overall intent to treat effect, as shown in Figures 10 and 11 in Appendix A. This is consistent with the findings on overall literacy and numeracy tests. Although the project did not have a significant impact on test scores *overall*, the project's impact was isolated to those who had been struggling most at the beginning of the program, particularly on basic literacy and numeracy skills. This focus on foundational skills for struggling students' improvement may explain why we do not observe any improvements overall compared to the control group.

FIGURE 5. Coefficients on Interaction Between IGATE-T Impact and Baseline Test Scores on Literacy Test Subtasks

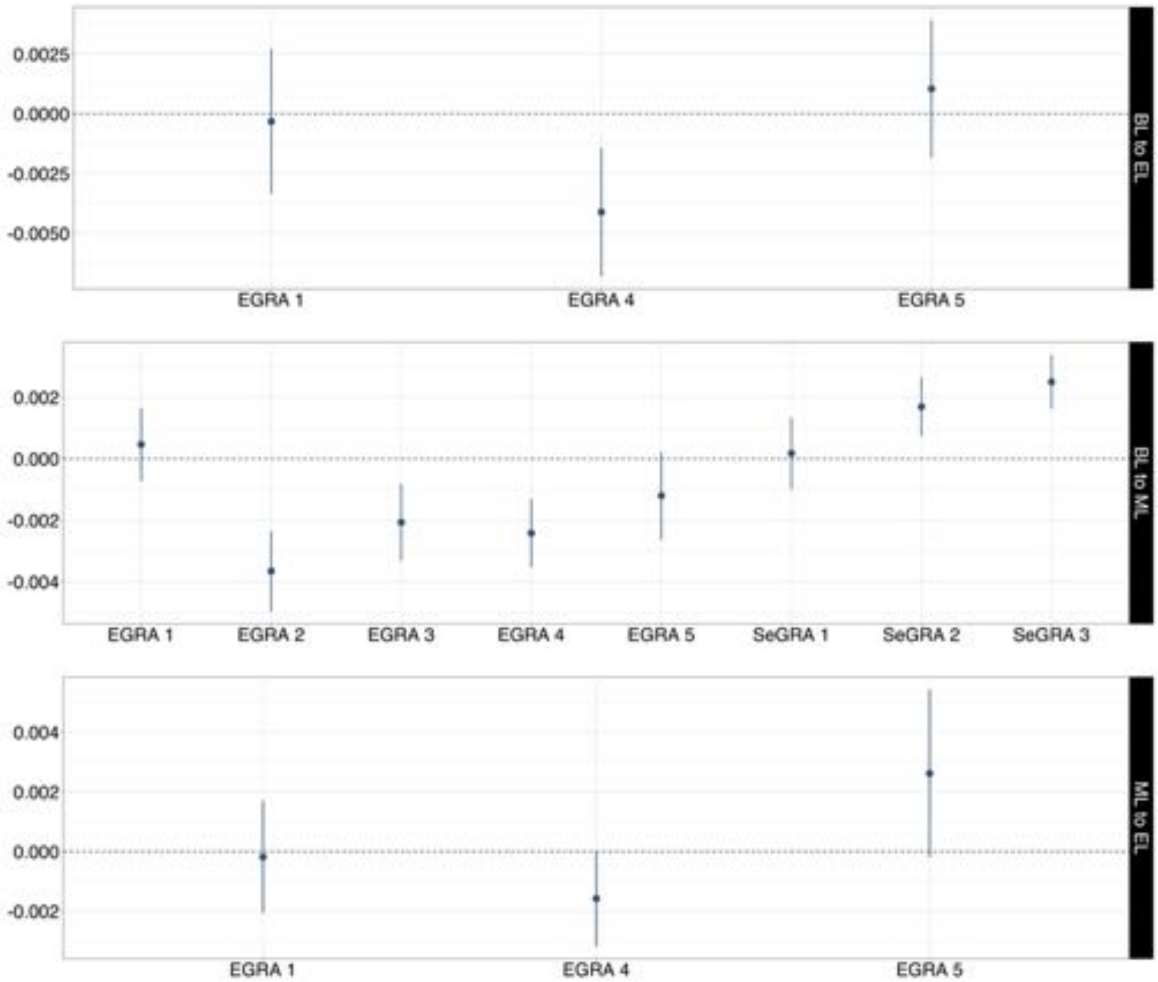


Figure note: The number of subtasks assessed at endline was reduced due to the project's budget constraints.

To examine the mechanisms through which IGATE-T has had an impact, Figure 7 shows the reported attitudes of teachers, headteachers, and officials from the Zimbabwe Ministry of Primary and Secondary Education associated with the question "Have the IGATE-T learning resources been effective?" At midline, the distribution of sentiment was positive on average. However, at endline there were consistently more favourable attitudes towards the learning materials provided by the IGATE-T program.[41] This indicates that more educators in the community were aware of the

[41]This difference is statistically significant with a p-value of 0.04 from the t-test testing the null hypothesis that there is no difference in means.

Figure 6. Coefficients on Interaction Between IGATE-T Impact and Baseline Test Scores on Numeracy Test Subtasks

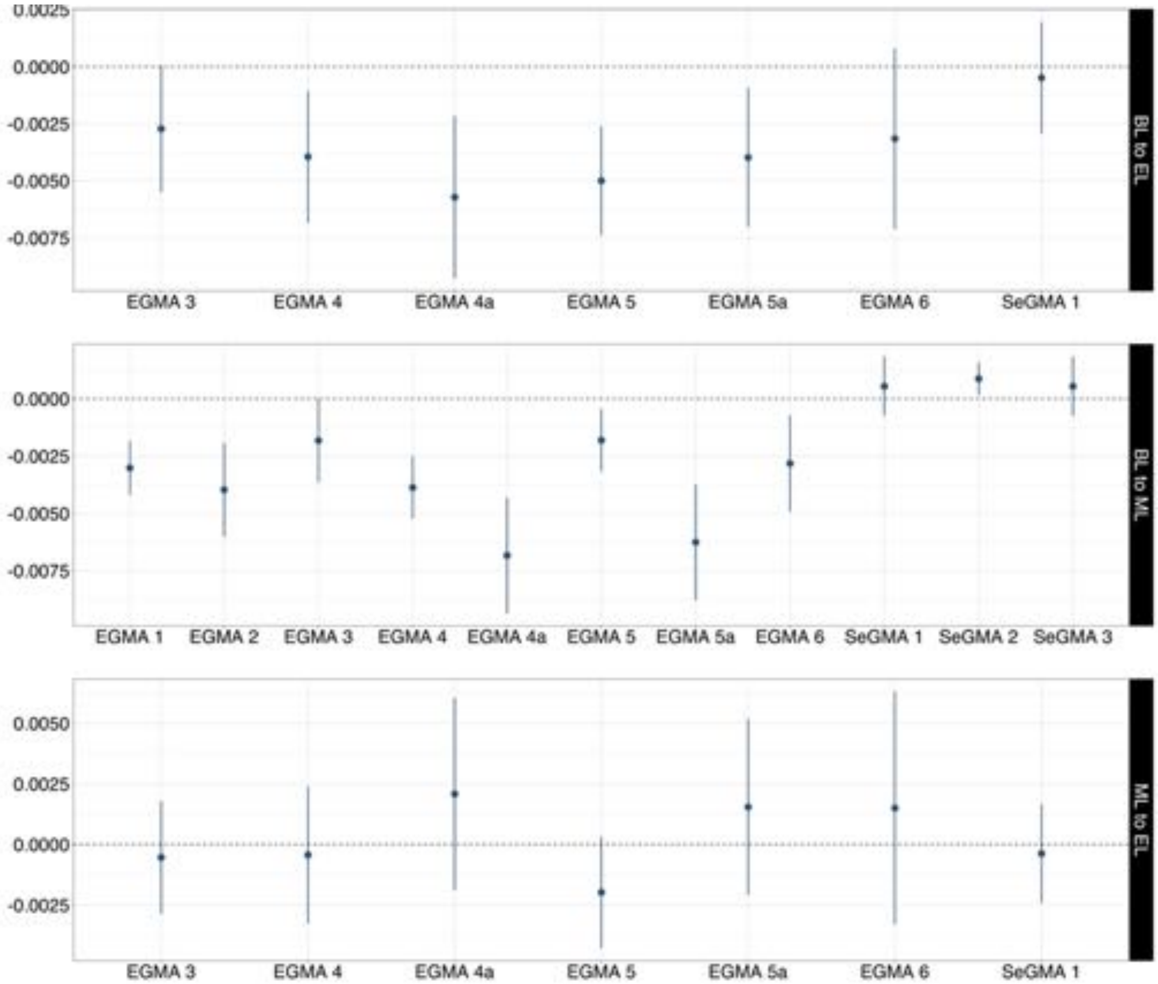
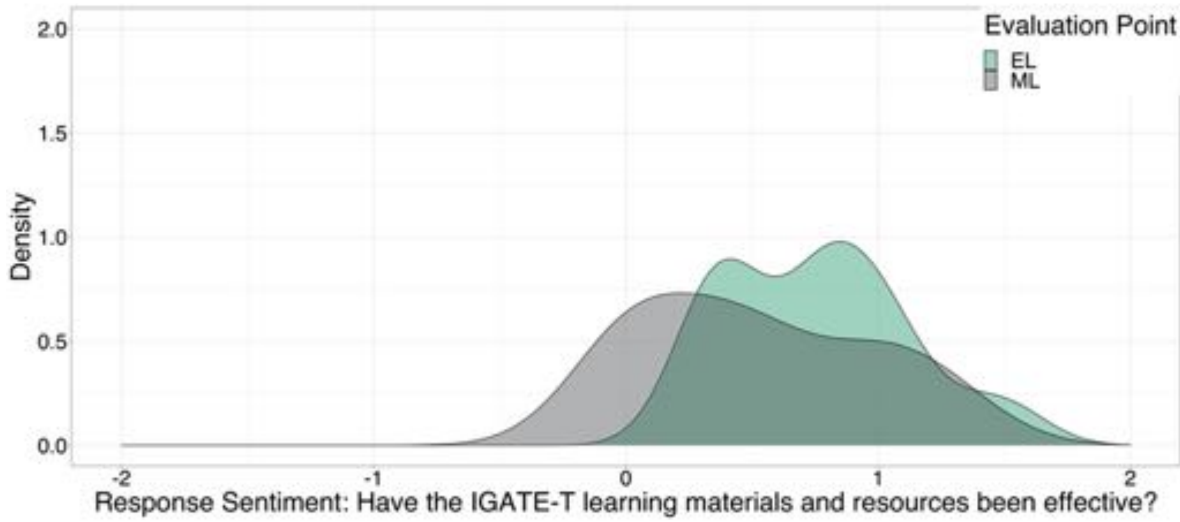


Figure note: The number of subtasks assessed at endline was reduced due to the project's budget constraints.

changes that were taking place, or that the communities found the interventions implemented post-midline—which included the Community Learning Circles—to be more effective in supporting learning.

Data from the qualitative interviews can be mined to specifically identify which parts of the IGATE-T interventions were identified as helpful in the period after midline, when educators had more positive attitudes towards the IGATE-T interventions. When asked about how the IGATE-T project has supported families and children during the last year of the project (and through the pandemic), community members and learners most often reported improved teaching practices, the community learning circles, and specific

Figure 7. Reported Sentiment Associated with IGATE-T Learning Resources



resources like books and reading cards—which had short stories and comprehension questions designed to various reading levels—were given out to the community learning circles during school closures. This is based on an analysis of the terms that respondents most frequently used in their answers, and on the associations between the terms they used.[42,43]

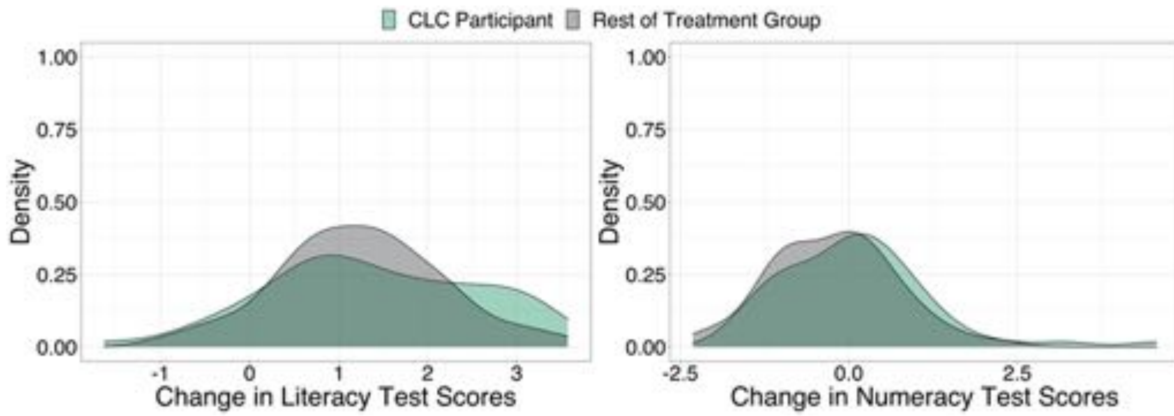
To examine whether this can be observed in the quantitative evidence, I look at the difference in test score improvements for individuals who had specifically participated in the community learning circles. This cannot be interpreted as causal since participation in the community learning circles was not automatic for those in treatment schools. However, when combined with the qualitative evidence, this analysis can explain the mechanisms that explain the project's role in affecting learning outcomes.

Figure 8 shows how test scores have changed for community learning circle participants since baseline, compared to members of the treatment group who had not participated in the community learning circles.[44] Compared to other members of the

[42]Accompanying figures can be found in Figures 12 and 13.
[43]These are supported by findings from content analysis of the focus group and key informant interview data, which found that community learning circles effectively provided access to learning opportunities that would have been otherwise unavailable to learners in IGATE-T communities during school closures and that educators and community members widely believed that these contributed to improvements in learning.
[44]The rest of the treatment group has been selected as the comparison to isolate for the marginal effect of the community learning circles, since the treatment group had access to other interventions that would make the control group less comparable for subgroup analysis involving the community learning circles.

treatment group, community learning circle participants are more likely to have done significantly better in both literacy and numeracy by endline.[45]

FIGURE 8. Changes in Literacy Test Scores by CLC Participation



This may be the result of the community's increased support for girls' education, since caregivers were one of the main barriers to students attending these centres.[46] However, it could also be attributable to increased awareness of teaching practices that can support education outcomes, which may have been used during the CLC implementation.

5.3. **Progression.** To estimate the average impact on dropout rates and repetition rates, Table 8 shows the results of the estimation of equation 2, with and without controls. I find no evidence that the project has had an economically or statistically significant impact on either progression outcome overall since the project began.

To understand why there may be no impact on progression *overall*, we should first examine the major barriers that prevent learners from progressing successfully through school and consider how these relate to the barriers that the IGATE-T project was designed to mitigate. Data from qualitative interviews can provide important insights into why the project may not have had a meaningful impact on overall dropout rates or grade repetition. The open-ended nature of the qualitative interviews makes it possible to identify the reasons that girls and community members cite for barriers to progression without making assumptions about what these reasons are ahead of time. When asked about the reasons that girls quit school, interviewees were most likely to report two

[45]The p-value from the t-test of null that the means equivalent is 0.090.
[46]Students who reported receiving enough support from their families to continue learning through the pandemic were 25.3 percentage points more likely to attend the community learning circles compared to other students in the treatment group.

Table 8. Impact of IGATE-T on Dropout and Repetition Rates

	Dropout		Grade Repetition	
	(1)	(2)	(3)	(4)
Progression by Endline				
Treatment	-0.00057	-0.0049	0.0025	0.0053
	(0.0259)	(0.0270)	(0.0281)	(0.0310)
Controls		✓		✓
Learners	1,228	1,228	1,228	1,228
Progression by Midline				
Treatment	0.00977	0.00565	0.02260	0.01246
	(0.0101)	(0.0091)	(0.0166)	(0.0191)
Controls		✓		✓
Students	1,228	1,228	1,228	1,228

Note: The table reports the marginal effect on the dropout and repetition rates. Cluster-robust standard errors are in parentheses. Standard errors are clustered at the school level. Controls include the learner's age and grade as well as baseline characteristics of the learner's district, household socioeconomic status (measured by the household's ability to pay for basic needs), and caregiver education level. * $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$

key challenges. The first is school fees, and a lack of resources or money to pay for school fees and levies.[47] As discussed above, school fees and levies are a significant barrier to education, particularly at the secondary school level when fees increase. This is a significant barrier to students completing their education, particularly for girls, since traditional gender norms may lead households to allocate resources towards male children over female children (Duflo et al., 2021).

The second is marriage and pregnancy. In Zimbabwe, the adolescent fertility rate (the number of births per 1,000 women ages 15-19) was 86 when IGATE-T began in 2017 (The World Bank, 2021).[48] The project did not specifically focus on reducing the incidence of early pregnancy through sexual and reproductive health measures, and the

[47]This is consistent with Snilstveit et al. (2015), which shows that limited household resources are a major barrier to education at all levels. In this review, the authors find that multi-component education interventions, like IGATE-T, rarely achieve widespread major education improvements without addressing these underlying resource constraints.

[48]The rate fell to 80 per 1000 girls by 2019. Updated figures are not available for the COVID-19 period at the time of writing. However, this is expected to have risen since 2019 as girls were more vulnerable during the COVID-19-related school closures (Chineka and Kurevakwesu, 2021; Dudzai and Wamara, 2021). The rate has consistently been lower than the rate in Sub-Saharan Africa and many other neighbouring countries since the early 1980s but is still substantially higher than the global average.

project has not had a significant impact on pregnancy rates overall.[49] However, IGATE-T's focus on changing community attitudes towards girls who have become pregnant or are already mothers should make it easier for girls to go back to school or pursue other informal education options.

Together, these findings indicate that financial resources and pregnancy are the largest barriers to students' progression outcomes. The project did not specifically address either of these barriers, which may explain why the project did not have a significant impact on dropout or repetition rates by either midline or endline. However, it did address general attitudes towards marginalized groups, specifically girls and struggling students. Table 9 shows the heterogeneous effects of the IGATE-T interventions on students who had been struggling in literacy and numeracy at baseline.

Table 9. Heterogeneous Effect of IGATE-T on Progression Outcomes

Progression by Endline	Dropout (1)	Grade Repetition (2)	Dropout (3)	Grade Repetition (4)
Treatment	0.039	0.0721	0.0159	0.0450
	(0.037)	(0.0458)	(0.0539)	(0.0672)
Treatment x BL Literacy	-0.00151	-0.00223**		
	(0.000957)	(0.00102)		
Treatment x BL Numeracy			-0.000480	-0.000943
			(0.000902)	(0.0011)
Controls	✓	✓	✓	✓
Students	1,228	1,228	1,228	1,228

Note: The table reports the marginal effect on the dropout and repetition rates. Cluster-robust standard errors are in parentheses. Standard errors are clustered at the school level. Controls include the learner's age and grade as well as baseline characteristics of the learner's district, household socioeconomic status (measured by the household's ability to pay for basic needs), and caregiver education level. * $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$

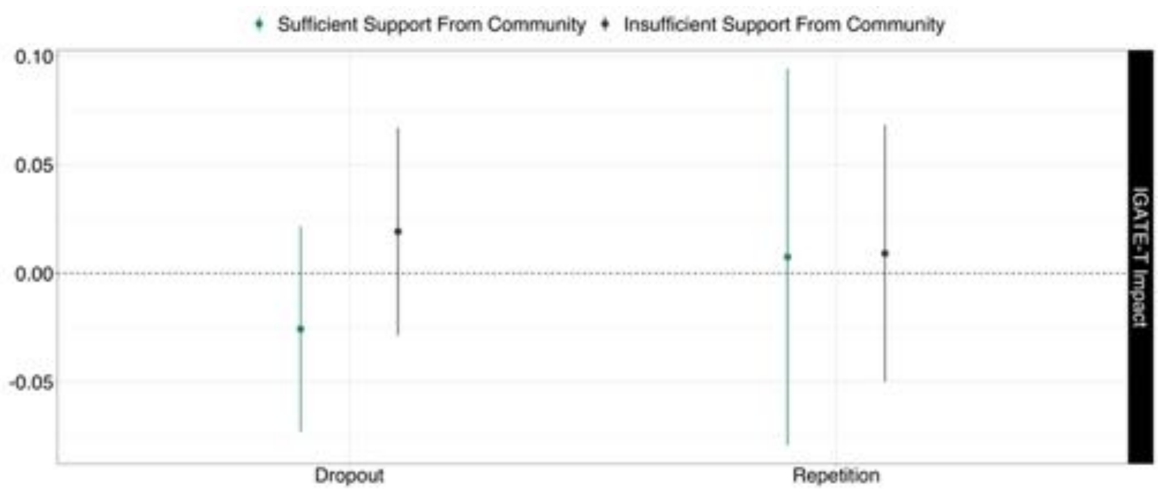
Table 9 shows that the project has also had some impact on progression outcomes for struggling students with repetition rates being 0.2 percentage points lower for every 1 percentage point worse a student did on baseline literacy tests. There is no heterogeneity in the effect for students who had been struggling with numeracy. However, when considered alongside the findings in table 7, this is further evidence that there are some heterogeneous treatment effects favouring struggling students.

To further examine the channels that explain this heterogeneity, figure 9 shows the differences in progression outcomes for students who reported receiving enough support

[49] The lack of significant results may be due to a limited number of cases where girls became pregnant by endline. By endline, 15 girls in the treatment group were mothers or were pregnant, compared with 21 in the control group. See Table 12 in Appendix A.

from their communities to continue their education during school closures, compared to those who did not receive this kind of support. The plot shows the treatment effects by support level with 95% confidence intervals. Although dropout rates are slightly lower for those reporting greater support from their communities, the IGATE-T interventions did not have a statistically significant impact on either progression outcome studied here when the sample is separated by the amount of support they reported receiving. However, these changes in support may not represent tangible changes in behaviours that are needed to support progression outcomes.

Figure 9. Progression by Community Support



The pandemic and the associated school closures introduced new barriers to education in these communities. With schools closed, most response efforts in Zimbabwe focused on providing online alternatives to learning, with some radio-based approaches as well. However, many students in IGATE-T communities come from households with limited access to the technologies online approaches rely on. The IGATE-T project responded to the crisis by establishing the community learning circles, which provided communities with low-tech resources such as reading cards and grade-specific workbooks, and trained community volunteers—including many school teachers and the project's established "community champions"—to facilitate informal teaching sessions. Student participation in the community learning circles depended on the support the students received from their caregivers and their communities to continue their education during the school closures, making their participation in the community learning circles an effective proxy for actionable community support.

In the discussion of the project's impact on test scores and learning, I showed that participants in community learning circles did significantly better on literacy and numeracy tests after participating in these groups. Table 10 examines the differences in the progression rates for community learning circle participants and non-participants in the treatment group using the same specification outlined in equation 2, with the treatment referring to the community learning circles.

Table 10. Marginal Effect of CLC on Progression Outcomes

	Dropout	Grade Repetition
Progression by Endline		
CLC Participation	-0.078**	-0.023
	(0.033)	(0.039)
Controls	✓	✓
Learners	614	614

Note: The table reports the marginal effect on the dropout and repetition rates. Cluster-robust standard errors are in parentheses. Standard errors are clustered at the school level. Controls include the learner's age and grade as well as baseline characteristics of the learner's district, household socioeconomic status (measured by the household's ability to pay for basic needs), and caregiver education level. * $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$

Community learning circle participants are over 7 percentage points less likely to drop out of school by endline, compared to the rest of the treatment group. As in the test scores results, this cannot be interpreted as causal since participation in the community learning circles was not automatic for those in treatment schools. However, when combined with the rest of the evidence presented thus far, this analysis may explain some mechanisms that define how the project is affecting progression outcomes. The community-based approach taken to implement the community learning circles implies that the types of households that were reached through the community learning circles varied slightly by district,[50] but there is a clear association between how much support students received from their families and their communities and how likely they were to attend the community learning circles.

5.4. **Possible Confounding Factor: Teaching Practices.** To convincingly attribute the changes in learning and progression to changes in community attitudes towards marginalized students, I examine another possible mechanism. Given the project's emphasis on providing teacher training, it is possible that the program's impact on the weakest students is being facilitated by changes in more general teaching practices

[50]See Appendix C for more details on differences in community learning circle participation by district.

themselves that would support struggling students. The emphasis on using participatory teaching methods may have led to significant improvements in learning for the weakest students.

As discussed in section 3.1, the learner survey asks students about the types of teaching practices their teacher used in lessons. Table 11 shows that the IGATE-T program did not specifically impact how often teachers used different participatory teaching methods that students were asked about. The table shows that the incidence of physical discipline fell by 3.2 percentage points. Students who were struggling with literacy or numeracy at baseline were also more likely to report physical punishments being used by their teachers at baseline. This may suggest that by increasing teachers' awareness of the challenges these students face, the project has impacted the types of discipline used in classrooms towards marginalized students.

Table 11. IGATE-T Impact on Teaching Practices

	Encourages Questions	Uses Resources	Frequent Absences	Uses Examples	Physical Punishments
EL-BL DiD	(1)	(2)	(3)	(4)	(5)
Treatment x Time	-0.00382	-0.00223	0.00475	-0.0111	-0.0319*
	(0.00445)	(0.00554)	(0.0172)	(0.00906)	(0.0183)
Controls	✓	✓	✓	✓	✓
Learners	1,228	1,228	1,228	1,228	1,228

Note: The table reports the marginal effect on the likelihood that students reported observing different teaching practices in class. Cluster-robust standard errors are in parentheses. Standard errors are clustered at the school level. Controls include the learner's age and grade as well as baseline characteristics of the learner's district, household socioeconomic status (measured by the household's ability to pay for basic needs), and caregiver education level. * $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$

Together, this could suggest that the program's impact on marginalized students is not being channelled through changes in specific teaching methods being used by teachers. However, earlier findings indicate that the teachers' practices are reported by community members as being a particularly important part of the project's contribution. Together, this is suggestive evidence that the changes in teacher *attitudes*, rather than specific teaching *methods* may contribute to the changes observed in education outcomes.

6. Discussion

This paper uses a mixed-methods study design involving a quasi-randomized field experiment and text mining analysis of qualitative data to understand the impact of interventions targeting community attitudes and teacher practices. Unlike other interventions targeting the education of marginalized groups in developing countries, there is little evidence on the efficacy of interventions targeting attitudes. This paper fills this gap by studying the impact of the IGATE-T interventions, which ran regular engagement sessions within communities to identify and discuss the barriers to education in their community, and established networks of community champions who worked with local leaders to encourage caregivers within their communities to adopt more supportive attitudes towards girls and marginalized students.

I find that community attitudes towards girls' education significantly improved after three and a half years of exposure to the program. By endline, community attitudes had improved by 0.403 SD towards girls' education in general, and by 0.63 SD towards adolescent mothers, who were traditionally a more marginalized group in most IGATE-T communities. This translated into a 20.9 percentage point increase in the number of students who reported receiving enough support to continue learning during the pandemic, compared to the control group.

There are several indicators that this change in attitudes led to meaningful behaviour changes within communities. After IGATE-T, girls were doing 31 fewer minutes of chores each day and students who reported receiving support from their communities were 25 percentage points more likely to have participated in community learning circles that provided informal education to students while schools were closed due to COVID-19. The heterogeneous treatment effects also show that these interventions were particularly effective for marginalized students, including girls and students struggling academically. Compared to students in the control group, IGATE-T led to 0.28 SD improvement in literacy for those who had been below the median at baseline, and for every one percentage point worse a student did on literacy tests at baseline, they were 0.2 percentage points less likely to repeat a grade.

These findings highlight the importance of community support in education interventions. For example, community support was a major barrier to the project's COVID-19 response, and the interventions were more effective when communities were supportive of these activities.

This research makes two important contributions to the literature. First, these results show that community attitudes can be influenced in a relatively short amount of time to become more supportive towards marginalized groups. Moreover, by allowing for heterogeneous treatment effects, this study shows that this type of intervention is also particularly effective in improving learning and progression outcomes for struggling students. The results of this study also find highly suggestive evidence that these changes in community attitudes led to improved learning and progression outcomes.

The second contribution is in the demonstration of how sentiment analysis and text mining methods can be used to measure changes in community attitudes. Though widespread in other disciplines, these tools (and qualitative data in general) are not commonly used in development economics or in economics more broadly (Starr, 2014). This study demonstrates how using qualitative evidence can provide important insights into the mechanisms along causal pathways, and that this can complement the attribution conclusions drawn from RCTs. If qualitative data was collected from control areas, these methods could also be used to make causal conclusions about changes in sentiment on specific topics.

References

Adukia, A. (2017). Sanitation and education. *American Economic Journal: Applied Economics 9*(2), 23–59.

Agarwal, A., B. Xie, I. Vovsha, O. Rambow, and R. J. Passonneau (2011). Sentiment analysis of Twitter data. In *Proceedings of the workshop on language in social media (LSM 2011)*, pp. 30–38.

Agarwal, M., V. Bahure, and S. Javadekar (2021). Age at marriage: Evidence from large scale education program.

Alesina, A., P. Giuliano, and N. Nunn (2013). On the origins of gender roles: Women and the plough. *The Quarterly Journal of Economics 128*(2), 469–530.

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.

Attanasio, O. P., C. Meghir, and A. Santiago (2012). Education choices in Mexico: using a structural model and a randomized experiment to evaluate Progresa. *The Review of Economic Studies 79*(1), 37–66.

Balahur, A., R. Steinberger, M. Kabadjov, V. Zavarella, E. Van Der Goot, M. Halkia, B. Pouliquen, and J. Belyaeva (2013). Sentiment analysis in the news. *arXiv preprint arXiv:1309.6202*.

Bandiera, O., N. Buehren, R. Burgess, M. Goldstein, S. Gulesci, I. Rasul, and M. Sulaiman (2020). Women's empowerment in action: Evidence from a randomized control trial in Africa. *American Economic Journal: Applied Economics 12*(1), 210–59.

Banerjee, A., R. Banerji, J. Berry, E. Duflo, H. Kannan, S. Mukherji, M. Shotland, and M. Walton (2016). Mainstreaming an effective intervention: Evidence from randomized evaluations of "Teaching at the Right Level" in India. *National Bureau of Economic Research*.

Barrera-Osorio, F., D. S. Blakeslee, M. Hoover, L. L. Linden, and D. Raju (2011). Expanding educational opportunities in remote parts of the world: evidence from a RCT of a public private partnership in Pakistan.

Bettinger, E. P. (2012). Paying to learn: The effect of financial incentives on elementary school test scores. *Review of Economics and Statistics 94*(3), 686–698.

Birdthistle, I., K. Dickson, M. Freeman, and L. Javidi (2011). What impact does the provision of separate toilets for girls at schools have on their primary and secondary school enrolment, attendance and completion? A systematic review of the evidence.

Social Science Research Unit, Institute of Education, University of London 6.

Buchmann, N., E. Field, R. Glennerster, S. Nazneen, S. Pimkina, and I. Sen (2017). Power vs money: Alternative approaches to reducing child marriage in Bangladesh, a randomized control trial. *Unpublished Manuscript*.

Buchmann, N., E. M. Field, R. Glennerster, S. Nazneen, and X. Y. Wang (2021). A signal to end child marriage: Theory and experimental evidence from Bangladesh. *National Bureau of Economic Research*.

Burde, D. (2004). Weak state, strong community? Promoting community participation in post-conflict countries. *Current Issues in Comparative Education 6*(2), 73–87.

Burde, D. and L. L. Linden (2013). Bringing education to Afghan girls: A randomized controlled trial of village-based schools. *American Economic Journal: Applied Economics 5*(3), 27–40.

Carvalho, S. and S. Hares (2020). More from our database on school closures: New education policies may be increasing educational inequality. *Center for Global Development*.

Chang, W., L. Diaz-Martin, A. Gopalan, E. Guarnieri, S. Jayachandran, and C. Walsh (2020). What works to enhance women's agency: Cross-cutting lessons from experimental and quasi-experimental studies. *J-PAL Working Paper*.

Chen, Z., X. Shi, W. Zhang, and L. Qu (2020). Understanding the complexity of teacher emotions from online forums: A computational text analysis approach. *Frontiers in Psychology 11*, 921.

Chineka, T. S. and W. Kurevakwesu (2021). Challenges for child welfare and development during the COVID-19 pandemic in Zimbabwe. *African Journal of Social Work 11*(4), 209–215.

Cotton, C., J. Nanowski, A. Nordstrom, and E. Richert (2020). Can community information campaigns improve girls' education? *Queen's Economics Department Working Paper*.

Cotton, C., A. Nordstrom, and Z. Robb (2021). Pedaling to prosperity: Understanding the impact of bicycles on education outcomes. *Queen's Economics Department Working Paper*.

Creswell, J. W. and C. N. Poth (2016). *Qualitative inquiry and research design: Choosing among five approaches*. Sage publications.

Cullen, J. B., S. D. Levitt, E. Robertson, and S. Sadoff (2013). What can be done to improve struggling high schools? *Journal of Economic Perspectives 27*(2), 133–52.

Dalal, S., B. Adlim, and M. Lesk (2019). How to measure relative bias in media coverage? *Significance 16*(5), 18–23.

Dhar, D., T. Jain, and S. Jayachandran (2019). Intergenerational transmission of gender attitudes: Evidence from india. *The Journal of Development Studies 55*(12), 2572–2592.

Dhar, D., T. Jain, and S. Jayachandran (2021). Reshaping adolescents' gender attitudes: Evidence from a school-based experiment in India. *National Bureau of Economic Research*.

Dudzai, C. and C. K. Wamara (2021). COVID-19 pandemic and the informal sector in Zimbabwe. *African Journal of Social Work 11*(4), 201–208.

Duflo, E., P. Dupas, and M. Kremer (2021). The impact of free secondary education: Experimental evidence from Ghana. *National Bureau of Economic Research*.

Gibbs, B. G. (2010). Reversing fortunes or content change? Gender gaps in math-related skill throughout childhood. *Social Science Research 39*(4), 540–569.

Gneezy, U., S. Meier, and P. Rey-Biel (2011). When and why incentives (don't) work to modify behavior. *Journal of Economic Perspectives 25*(4), 191–210.

Hanushek, E. A. and S. G. Rivkin (2012). The distribution of teacher quality and implications for policy. *Annu. Rev. Econ. 4*(1), 131–157.

Hastie, T., J. Qian, and K. Tay (2016). An introduction to glmnet.

Hastie, T., R. Tibshirani, and J. Friedman (2009). *The Elements of Statistical Learning* (Second ed.). Springer.

Hu, M. and B. Liu (2004). Mining and summarizing customer reviews. In *Proceedings of the Tenth ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, pp. 168–177.

Hu, Y. and W. Li (2011). Document sentiment classification by exploring description model of topical terms. *Computer Speech & Language 25*(2), 386–403.

Ikoro, V., M. Sharmina, K. Malik, and R. Batista-Navarro (2018). Analyzing sentiments expressed on Twitter by UK energy company consumers. In *2018 Fifth International Conference on Social Networks Analysis, Management and Security (SNAMS)*, pp. 95–98. IEEE.

Jacobs, A. M., B. Herrmann, G. Lauer, J. Lüdtke, and S. Schroeder (2020). Sentiment analysis of children and youth literature: Is there a Pollyanna effect? *Frontiers in Psychology 11*, 2310.

James, G., D. Witten, T. Hastie, and R. Tibshirani (2013). *An Introduction to Statistical Learning with Applications in R*. Springer.

Jayachandran, S. (2015). The roots of gender inequality in developing countries. *economics 7*(1), 63–88.

Jockers, M. (2017). Package 'syuzhet'. *URL: https://cran. r-project. org/web/packages/syuzhet*.

Karlgren, J., M. Sahlgren, F. Olsson, F. Espinoza, and O. Hamfors (2012). Usefulness of sentiment analysis. In *European Conference on Information Retrieval*, pp. 426–435. Springer.

Kennedy, A. and D. Inkpen (2006). Sentiment classification of movie reviews using contextual valence shifters. *Computational Intelligence 22*(2), 110–125.

Klasen, S. (2018). The impact of gender inequality on economic performance in developing countries. *Annual Review of Resource Economics 10*, 279–298.

Kravdal, Ø. (2002). Education and fertility in Sub-Saharan Africa: Individual and community effects. *Demography 39*(2), 233–250.

Kremer, M., E. Miguel, and R. Thornton (2009). Incentives to learn. *The Review of Economics and Statistics 91*(3), 437–456.

Leahey, E. and G. Guo (2001). Gender differences in mathematical trajectories. *Social Forces 80*(2), 713–732.

Lennox, R. J., D. Veríssimo, W. M. Twardek, C. R. Davis, and I. Jarić (2020). Sentiment analysis as a measure of conservation culture in scientific literature. *Conservation Biology 34*(2), 462–471.

Lipowsky, F., K. Rakoczy, C. Pauli, K. Reusser, and E. Klieme (2007). Gleicher unterricht-gleiche chancen für alle? die verteilung von schülerbeiträgen im klassenunterricht. *Unterrichtswissenschaft 35*(2), 125–147.

Liu, B. et al. (2010). Sentiment analysis and subjectivity. *Handbook of Natural Language Processing 2*(2010), 627–666.

Marks, G. N. (2008). Accounting for the gender gaps in student performance in reading and mathematics: Evidence from 31 countries. *Oxford Review of Education 34*(1), 89–109.

Mayring, P. (2004). Qualitative content analysis. *A Companion to Qualitative Research 1*(2), 159–176.

Ministry of Primary and Secondary Education (May 2020). 2019 Primary and secondary education statistics report. Technical report, Government of Zimbabwe.

Moestue, H. and S. Huttly (2008). Adult education and child nutrition: The role of family and community. *Journal of Epidemiology & Community Health 62*(2), 153–159.

Mohammad, S. M. (2012). From once upon a time to happily ever after: Tracking emotions in mail and books. *Decision Support Systems 53*(4), 730–741.

Mohammad, S. M. and P. D. Turney (2013). Crowdsourcing a word-emotion association lexicon. *Computational Intelligence 29*(3), 436–465.

Moursund, A. and Ø. Kravdal (2003). Individual and community effects of women's education and autonomy on contraceptive use in India. *Population Studies 57*(3), 285–301.

Musa Kika (June 2019). Age of consent, sexual intercourse with young persons and access to sexual and reproductive health care in Zimbabwe. Technical report, Justice for Children.

Naldi, M. (2019). A review of sentiment computation methods with R packages. *arXiv preprint arXiv:1901.08319*.

Neviarouskaya, A., H. Prendinger, and M. Ishizuka (2010). Recognition of affect, judgment, and appreciation in text. In *Proceedings of the 23rd International Conference on Computational Linguistics (Coling 2010)*, pp. 806–814.

Nielsen, F. Å. (2011, mar). AFINN.

Nillesen, E., M. Grimm, M. Goedhuys, A.-K. Reitmann, and A. Meysonnat (2021). On the malleability of gender attitudes: Evidence from implicit and explicit measures in tunisia. *World Development 138*, 105263.

Nordstrom, A. (2021). What affects the education outcomes of marginalized students? A machine learning approach to inform program design. *Queen's Economics Department Working Paper*.

Nordstrom, A. and C. Cotton (2021). Impact of a severe drought on education: More schooling does not imply more learning. *Available at SSRN 3601834*.

Rinker, T. (2017). Package 'sentimentr'. *Retrieved 8*, 31.

Rui, H., Y. Liu, and A. Whinston (2013). Whose and what chatter matters? The effect of tweets on movie sales. *Decision support systems 55*(4), 863–870.

Sailors, M. and M. Samati (2014). Community mobilization: Going beyond to support the implementation of a complementary reading program in malawi. *63rd yearbook of the Literacy Research Association*, 158–171.

Sebring, P. B., E. Allensworth, A. S. Bryk, J. Q. Easton, and S. Luppescu (2006). The essential supports for school improvement. research report. *Consortium on Chicago School Research*.

Seguino, S. (2011). Help or hindrance? religion's impact on gender inequality in attitudes and outcomes. *World Development 39*(8), 1308–1321.

Silge, J. and D. Robinson (2017). *Text mining with R: A tidy approach.* O'Reilly Media, Inc.

Singh, S. and R. Samara (1996). Early marriage among women in developing countries. *International Family Planning Perspectives*, 148–175.

Snilstveit, B., J. Stevenson, D. Phillips, M. Vojtkova, E. Gallagher, T. Schmidt, H. Jobse, M. Geelen, and M. Grazia-Pastorello (2015). Interventions for improving learning outcomes and access to education in low-and middle-income countries: A systematic review. *International Initiative for Impact Evaluation Systematic Reviews 24*.

Solovyev, V., M. Solnyshkina, E. Gafiyatova, D. McNamara, and V. Ivanov (2019). Sentiment in academic texts. In *2019 24th Conference of Open Innovations Association (FRUCT)*, pp. 408–414. IEEE.

Starr, M. A. (2014). Qualitative and mixed-methods research in economics: Surprising growth, promising future. *Journal of Economic Surveys 28*(2), 238–264.

Surridge, M., C. Chiroro, N. Marimo, T. Kureya, G. Zinumwe, T. Mukupe, K. Dombojena, and R. Roland (2020). Longitudinal study into dropout and survival in Zimbabwean schools final report. *Centre for International Development and Training*.

The World Bank (2021). Adolescent fertility rate (births per 1,000 women ages 15-10).

Tibshirani, R. (1996). Regression shrinkage and selection via the lasso. *Journal of the Royal Statistical Society. Series B (Methodological)*, 267–288.

White, H. (2013). The use of mixed methods in randomized control trials. *New Directions for Evaluation 2013*(138), 61–73.

Wilson, T., J. Wiebe, and P. Hoffmann (2005). Recognizing contextual polarity in phrase-level sentiment analysis. In *Proceedings of human language technology conference and conference on empirical methods in natural language processing*, pp. 347–354.

Wilson, T. A. (2008). *Fine-grained subjectivity and sentiment analysis: Recognizing the intensity, polarity, and attitudes of private states.* University of Pittsburgh.

Xianghua, F., L. Guo, G. Yanyan, and W. Zhiqiang (2013). Multi-aspect sentiment analysis for chinese online social reviews based on topic modeling and hownet lexicon. *Knowledge-Based Systems 37*, 186–195.

Xu, G., Z. Yu, Z. Chen, X. Qiu, and H. Yao (2019). Sensitive information topics-based sentiment analysis method for big data. *IEEE Access 7*, 96177–96190.

Yu, H., J. Shang, M. Hsu, M. Castellanos, and J. Han (2016). Data-driven contextual valence shifter quantification for multi-theme sentiment analysis. In *Proceedings of the 25th ACM International Conference on Information and Knowledge Management*, pp. 939–948.

Yu, L.-C., J.-L. Wu, P.-C. Chang, and H.-S. Chu (2013). Using a contextual entropy model to expand emotion words and their intensity for the sentiment classification of stock market news. *Knowledge-Based Systems 41*, 89–97.

Zou, H. and T. Hastie (2005). Regularization and variable selection via the elastic net. *Journal of the Royal Statistical Society: Series B (Statistical Methodology) 67*(2), 301–320.

Appendix A. Additional Tables and Figures

Figure 10. IGATE-T Impact on Literacy Test Subtasks

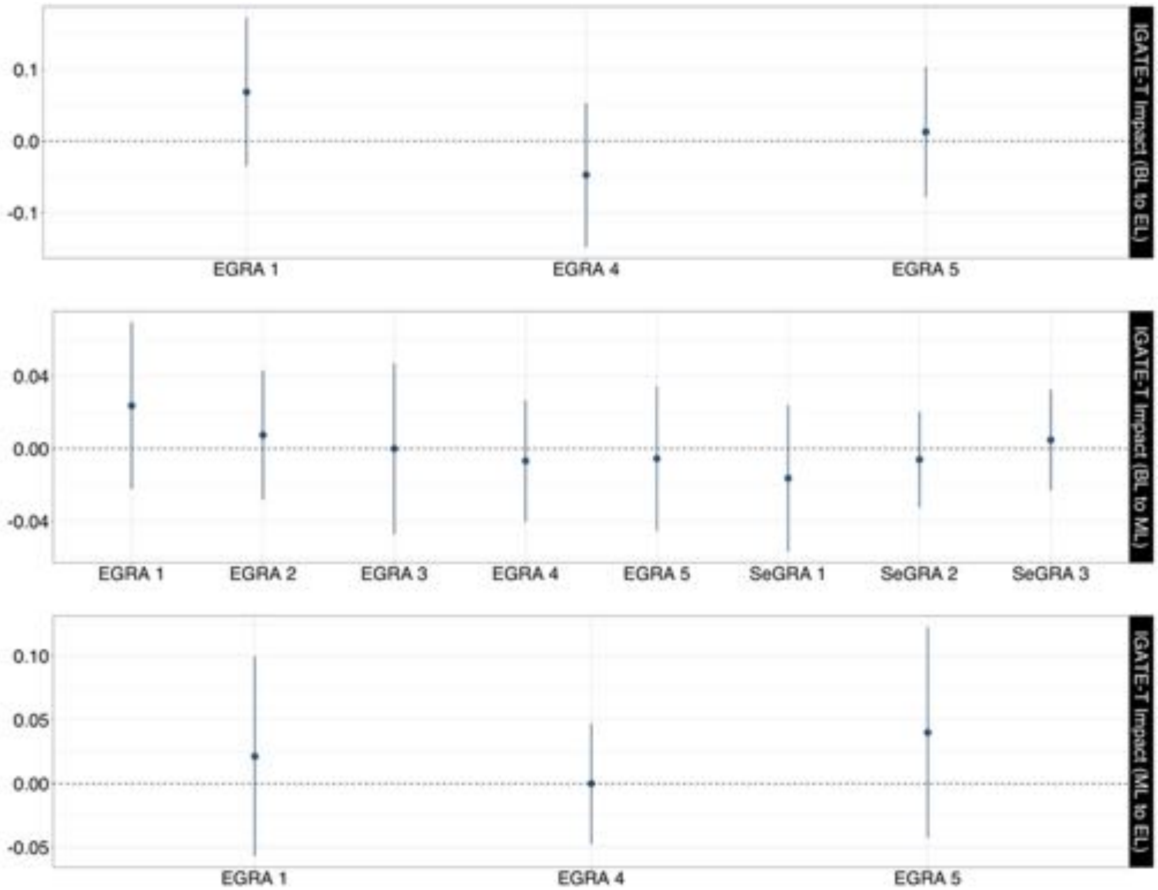


Figure note: The plot shows the treatment effects by subtask with 95% confidence intervals. The number of subtasks assessed at endline was reduced due to the project's budget constraints. The IGATE-T interventions did not have a significant impact on any of the subtasks that were included in the literacy assessments. This is consistent with the overall findings that find the IGATE-T intervention has had no impact on overall learning outcomes, before accounting for heterogeneous treatment effects.

Figure 11. IGATE-T Impact on Numeracy Test Subtasks

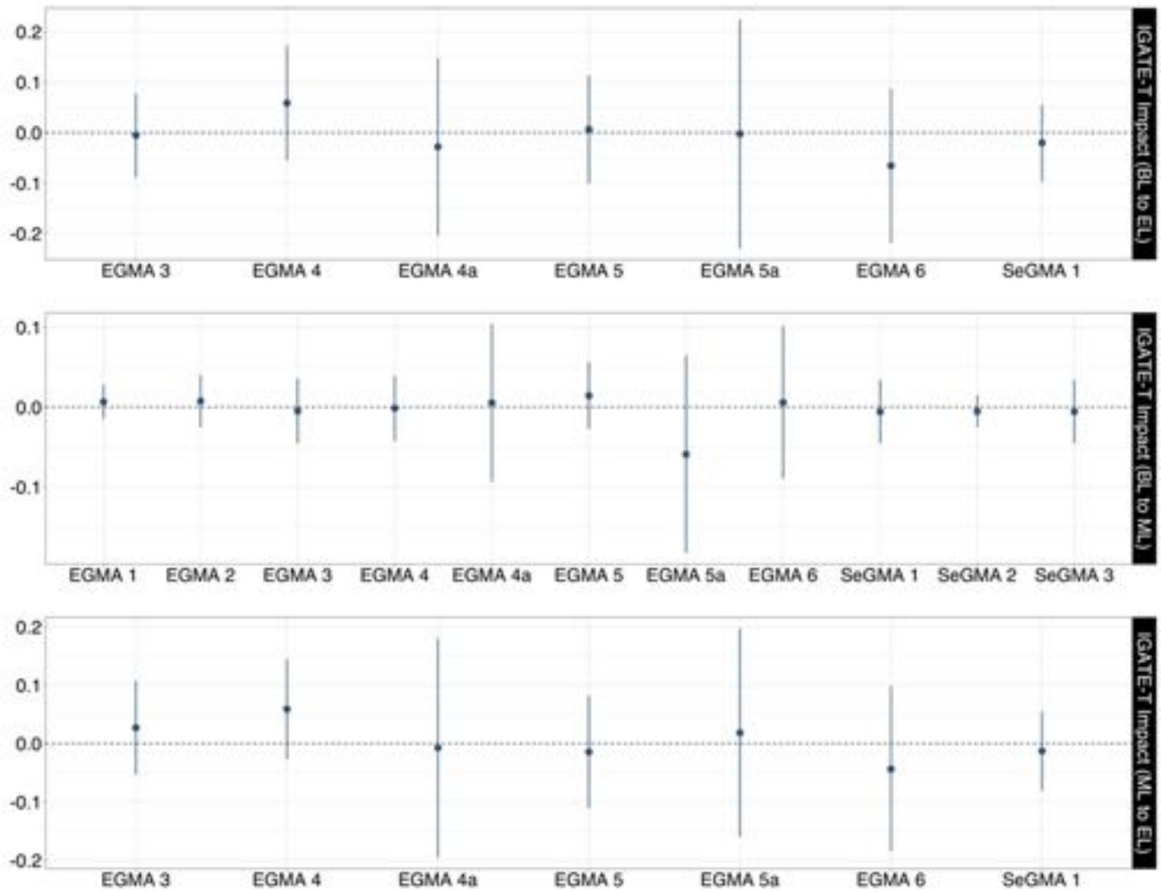


Figure note: The number of subtasks assessed at endline was reduced due to the project's budget constraints. The IGATE-T interventions did not have a significant impact on any of the subtasks that were included in the numeracy assessments. This is consistent with the overall findings that find the IGATE-T intervention has had no impact on overall learning outcomes, before accounting for heterogeneous treatment effects.

FIGURE 12. Most Common Terms From Responses to "How has IGATE supported families and children during the last year and throughout the pandemic?"

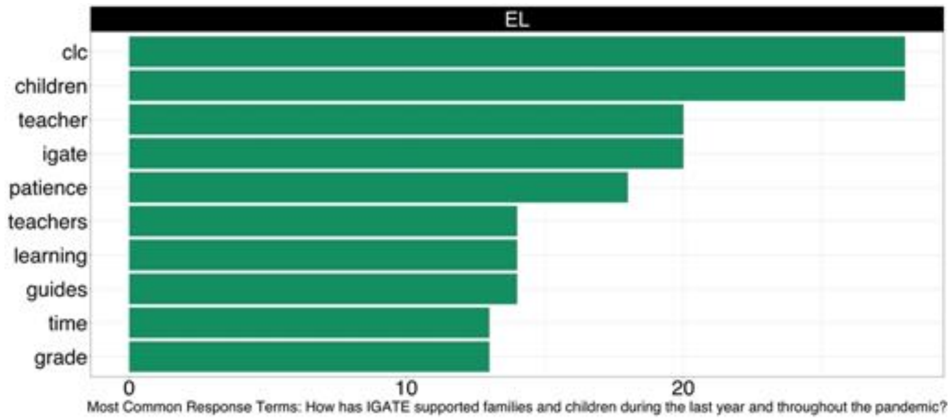
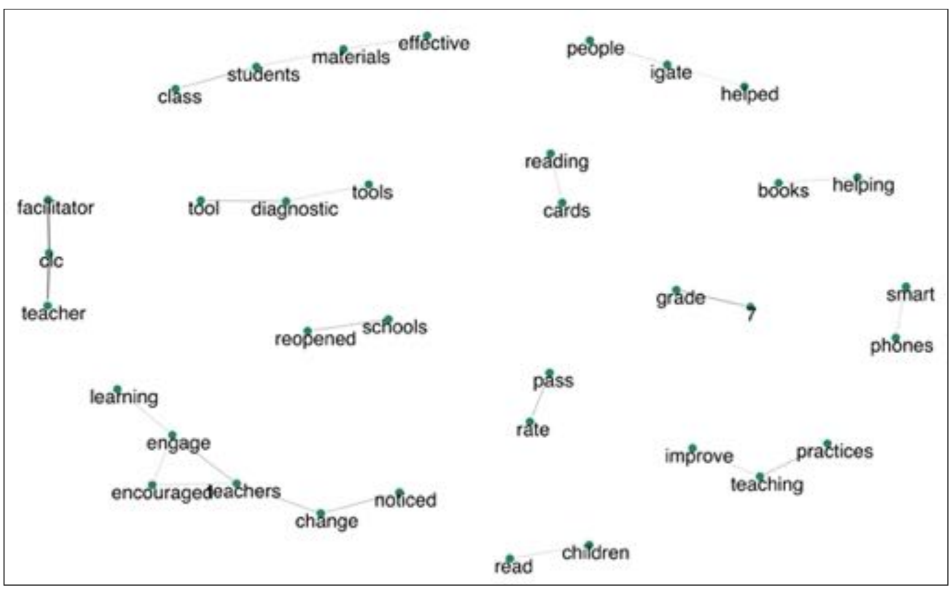


FIGURE 13. Most Common Term Associations From Responses to "How has IGATE supported families and children during the last year and throughout the pandemic?"



Figures 13 and 12 present the most common terms referenced in responses to questions about how IGATE-T has supported families and children during the pandemic, and the most common term associations in these responses, respectively. When asked about how the IGATE-T project has supported families and children during the last year of the project (through the pandemic), community members and learners most often reported improved teaching practices (see “teachers” and “patience” in Figure 13 and links between "teachers," "encouraged," "learning," and "engage," and associations between "improve," "teaching," and "practices," among others), the community learning circles, and specific resources like books and reading cards—which had short stories and comprehension questions designed to various reading levels—were given out to the community learning circles during school closure.

TABLE 12. Impact of IGATE-T on Adolescent Pregnancy

	Pregnant or Mother
By Endline	
CLC Participation	-0.014
	(0.011)
Controls	✓
Learners	1,060
By Midline	
CLC Participation	0.038
	(0.024)
Controls	✓
Students	1,060

Note: The table reports the marginal effect on the rates of adolescent pregnancy/motherhood, compared to the control group. Cluster-robust standard errors are in parentheses. Standard errors are clustered at the school level. Controls include the learner's age and grade as well as baseline characteristics of the learner's district, household socioeconomic status (measured by the household's ability to pay for basic needs), and caregiver education level. * $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$

Table 12 shows that the project has not had a statistically significant impact on rates of adolescent pregnancy or motherhood. The number of girls who became pregnant in the control group is higher than in the treatment group. However, the relatively small number of pregnancies limits the ability to comment on whether this difference is statistically significant.

Appendix B. Analysis of Attrition Rates

In this section, I examine the characteristics of those who attrited from the sample compared to those who have not attrited from the sample. Table 13 shows that being in a higher grade at baseline is associated with a higher probability of attriting from the sample by endline. This would be expected if learners who drop out of school are more likely to have attrited from the sample, since older students are also more likely to have dropped out of school. As discussed earlier, dropout rates for girls begin to climb as girls reach upper secondary school. By endline, girls who were in grades 6 and 7 (and were also, therefore, slightly older) are either in or approaching upper secondary school grades by endline, and would be more likely to drop out of school.

Table 13. Baseline Summary Statistics by Attrition Status

	Recontacted	**Attrited**	**Difference**
Grade	4.786	5.146	-0.360***
Age	10.755	11.165	-0.411***
Female	0.137	0.135	0.002
Pregnant or parent	0.002	0.002	0.000
Disability	0.092	0.101	-0.009
Lives without parent	0.281	0.340	-0.058
Orphan	0.152	0.198	-0.046***
Daily chores (hrs)	1.625	1.797	-0.172
HH Experiences Hunger	0.372	0.361	0.011
PCG has no education	0.100	0.093	0.007
Apostolic	0.306	0.280	0.027
Safe Commute	0.216	0.222	-0.006
Teacher frequently absent	0.276	0.254	0.021
No water at school	0.238	0.254	-0.016
Literacy Test Score	36.573	36.679	-0.106
Numeracy Test Score	56.664	56.255	0.409
YLI Score	55.012	55.187	-0.174
Attendance (/22 days)	19.096	19.143	-0.046
Observations	1228	491	737

Note: * $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$

When we consider the possible reasons for why students were not recontacted at endline, 76% of students had moved out of the area and were outside a reasonable distance to contact.[51] In addition to the 76% who had relocated, 7% could not be located, and 17% were unavailable during three separate contact attempts. This is consistent with the notion that older students are also more likely to have attrited, since these students are also more likely to leave the rural areas where they attended school if they have dropped out.

To further examine which (if any) baseline characteristics are systematically associated with attrition rates, I use two regularization methods to identify the characteristics—out of over 1,500 possible observable baseline characteristics in the data—that are the most important predictors of attrition. These regularization techniques allow me to do a more comprehensive study of the relationships between attrition and the over 1,500 observable characteristics available for each student at baseline, to conclude that apart from age—which is a well-established predictor of school dropouts—there are very few characteristics associated with attrition rates. Furthermore, none of these characteristics are associated with treatment status, which suggests that attrition should not bias the findings of this study.

The first, the "least absolute shrinkage and selection" (lasso) technique is a popular shrinkage method that reduces the number of predictors in a standard regression model. First introduced by Tibshirani (1996) as an alternative to ridge regression, subset selection, and OLS, the lasso technique involves shrinking predictors in a model to improve the model's interpretability and prediction accuracy by shrinking some of the parameters down to zero.

While lasso does improve the interpretability of linear models, the method tends to over-regularize, which may limit a model's explanatory power (Zou and Hastie, 2005; Hastie et al., 2009). Introduced by Zou and Hastie (2005), elastic net offers a less extreme alternative to lasso while still offering greater interpretability than other regularization alternatives such as ridge regression. For a binary response outcome variable (such as attrition status) elastic net parameters are estimated by solving the following objective function:

$$\min_{(\beta_0,\beta)\in\mathbb{R}^{p+1}} -\left[\frac{1}{N}\sum_{i=1}^{N} att_i\cdot\left(\beta_0+x_i^T\beta\right)-\log\left(1+e^{\left(\beta_0+x_i^T\beta\right)}\right)\right]+\lambda\left[(1-\alpha)\|\beta\|_2^2/2+\alpha\|\beta\|_1\right]$$

[51] When students could not be recontacted at the schools at either midline or endline, enumerators asked headteachers and caregivers—and neighbours, if the caregiver had previously consented to have neighbours be contacted during follow-up interviews—about the student's whereabouts to record this information.

where λ defines how much each parameter estimates will be reduced, and is typically selected using cross-validation, a common method used to select the optimal level of model complexity to achieve the best predictive power (as measured by the mean square error in the test set) (James et al., 2013; Hastie et al., 2016). att is a binary indicator for attrition, and β is a matrix of coefficients for the matrix of x of all the observable baseline characteristics. As λ increases, more parameters are reduced to zero. The elastic net penalty term allows for a combination of the penalty in lasso and the penalty in ridge regression, with the extreme cases where $\alpha = 1$ reducing to the standard lasso problem and $\alpha = 0$ reducing to the standard ridge problem. This is done to account for some of the correlation between groups of variables that are strongly correlated since neither ridge regression nor lasso is perfect at dealing with groups of highly correlated variables. Hastie et al. (2009) note that this is useful in the fields of genomics and proteomics, where problems often involve more parameters than observations and these predictors are highly correlated. These problems are also common in the social sciences and international development, which is why this method has been included in this application.

The left panel of Figure 14, shows the error associated with different levels of λ from cross-validation which was selected using cross-validation to minimize the mean squared error in the test set's predictions for attrition. The parameters associated with this optimal value of λ is shown in the right-hand panel, where the penalty term is indicated by the dotted line where $log(\lambda)$ is equal to -3.2. Note that at the optimal penalty term, all parameters (not including the intercept) have been reduced to zero, leaving no non-zero parameters to predict attrition. This would strongly suggest that there are very few systematic relationships between attrition and observable characteristics. Given the number of observable characteristics within this dataset, this would indicate that there are no systematic attrition patterns that would affect the interpretation of the main results.

This is largely consistent with what is observed when I relax the penalty terms in lasso to use elastic net. The left panel of Figure 15 shows the error associated with different levels of λ from cross-validation, at the optimal level of α, which was selected using two-dimensional cross-validation to minimize the mean squared error in the test set's predictions for attrition. The parameters associated with these values of λ and α are shown in the right-hand panel, where the optimal combination of penalty terms is indicated by the dotted line where $log(\lambda)$ is equal to -1.2.

Figure 14. Lasso Penalty and Coefficients

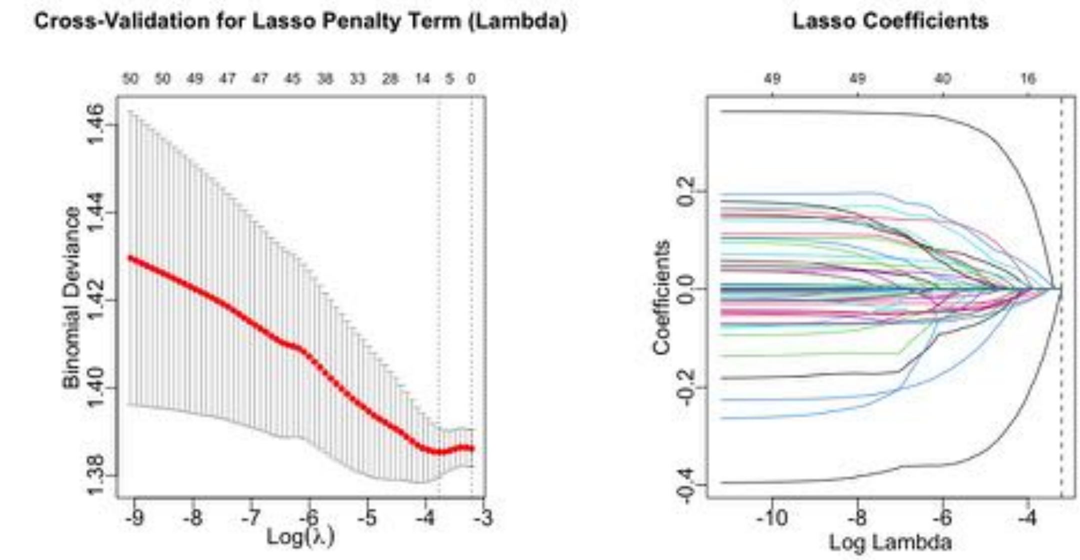


Figure 15. Elastic Net Penalty and Coefficients

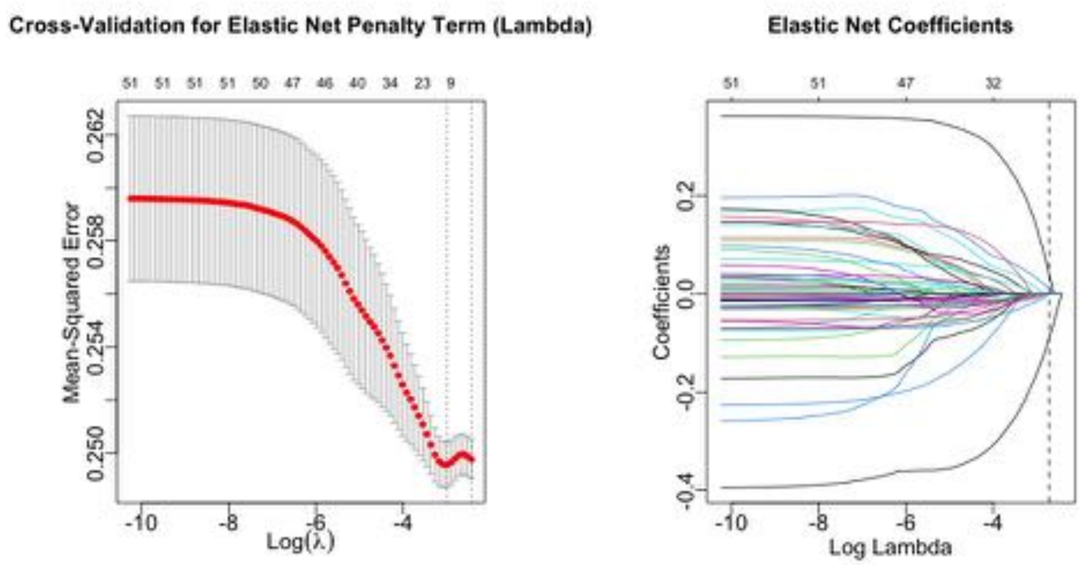


Table 14 shows the parameter estimates associated with an OLS regression that includes the non-zero parameters identified using elastic net. Notably, grade, district, and head of household education indicator variables have all been included in the elastic net specification. Again, I find that older students are more likely to attrite from the sample, as are students from the Insiza district. None of these characteristics are related to treatment status, which supports the conclusion that attrition rates are not systematically biasing the main results covered in this paper.

TABLE 14. Characteristics Predicting Attrition

Baseline Characteristic	**Elastic Net**
Grade 4	0.104**
	0.0338
Grade 5	0.0842**
	0.034
Grade 6	0.186***
	(0.0335)
Grade 7	0.190***
	(0.0341)
Head of Household has Completed Primary School	-0.095***
	(0.0248)
District Indicators	
- Insiza	0.132***
	0.0329
- Mangwe	-0.0738**
	0.0299
- Mberengwa	-0.0317
	(0.0280)
Mean Squared Error (Test Set)	0.251
Total Non-Zero Parameters (Not including intercept)	8
Students	1,719

Note: The table reports the marginal effect on the probability a student attrited from the sample. Robust standard errors are in parentheses. The variables included in the regression have been selected using regularized probit regression paths to minimize the test mean square error. α and λ were selected using two-dimensional cross-validation. * $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$

Appendix C. Community Learning Circle Participation Differences

Community learning circles were implemented slightly differently in each of the four geographic districts where the evaluation took place. The main difference was how the circles were targeted to caregivers with varying levels of education. In particular, learning circles in Chivi, Insiza, and Mangwe were promoted to caregivers who had not completed any secondary school. In these districts, the average participation rate in the learning circles was higher than in Mberengwa, where caregiver education levels were higher, as shown in Table 15.

Table 15. Caregiver Education Levels by District Amongst Community Learning Circle Participants

Caregiver Education Level	**Chivi**	**Insiza**	**Mangwe**	**Mberengwa**
None	27%	23%	46%	24%
Primary	42%	55%	47%	27%
Lower Secondary	20%	19%	4%	29%
Upper Secondary	8%	0%	4%	16%
University	2%	3%	0%	4%
Students	171	39	75	61
CLC Participation Rate	*41%*	*52%*	*49%*	*37%*

I find some evidence that the community learning circles were most effective in improving progression rates for students whose caregivers have not completed any secondary school. This is consistent with the broader findings surrounding community attitudes and the barriers to progression. Coming from a household with lower levels of caregiver education is associated with an increased risk of dropping out of school or repeating a grade. Before the project began, learners from households with lower levels of caregiver education were also significantly less likely to report receiving support to stay in school from their families. This suggests that by tailoring the community learning circles to address these barriers, progression outcomes for these students have improved.

To further examine the mechanisms through which the IGATE-T project and the community learning circles affected students, an interaction term between the community learning circle indicator and an indicator for whether or not a learner's caregiver has completed at least any secondary education has been added to the specification used to estimate Table 10. This added interaction term is shown in Table 16. While the community learning circles do not appear to have been affecting the dropout rates through this channel,[52] Column (2) shows that the parental education level does predict the community learning circle's effect on repetition rates. Specifically, students who have a caregiver who has not completed any secondary school and who attend the community learning circles are 12.9 percentage points less likely to repeat a grade.

Table 16. Heterogeneous Effect of CLC on Progression Outcomes (including interaction with caregiver education)

	Dropout	Grade Repetition
Progression by Endline		
CLC Participation	-0.066*	0.0092
	(0.038)	(0.044)
CLC Participation x Uneducated Caregiver	-0.030	-0.129**
	(0.050)	(0.059)
Controls	✓	✓
Learners	614	614

Note: The table reports the marginal effect on the dropout and repetition rates. Cluster-robust standard errors are in parentheses. Standard errors are clustered at the school level. Controls include the learner's age and grade as well as baseline characteristics of the learner's district, household socioeconomic status (measured by the household's ability to pay for basic needs), and caregiver education level. * $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$

[52]This may be a consequence of the relatively small number of learners who have dropped out of school and meet the restrictions used for this subgroup analysis. Only 9 students who participated in the community learning circles and have dropped out of school by endline, which significantly limits the power of this analysis.