CENTER FOR GLOBAL DEVELOPMENT

The Educational Impacts of Cash Transfers in Tanzania

David K. Evans, Charles Gale, and Katrina Kosec

Abstract

Cash transfers boost educational outcomes for poor children on average, but which aspects of educational performance are most responsive and which poor children benefit the most? This study examines the educational impacts of cash transfers, drawing on a randomized, community implemented conditional cash transfer program targeted to poor households in Tanzania. On average, being assigned to receive transfers significantly improves children's likelihood of having ever attended school (by between 4 and 5 percentage points), with suggestive evidence that this is driven by more age-appropriate enrollment for the youngest children. However, school attendance and primary school completion remain unaffected on average. Girls and boys benefit similarly, and only students with stronger initial educational performance experience increases in primary completion rates.

This paper was first published in December 2020. The original version is available at https://www.cgdev.org/sites/default/files/education-impacts-cash-transfers-children-multiple-indicators-vulnerability.pdf

KEYWORDS
cash transfers, education, Tanzania, poverty

JEL CODES
C93, I21, O12

WORKING PAPER 563 • JULY 2022

The Educational Impacts of Cash Transfers in Tanzania

David K. Evans
Center for Global Development; devans@cgdev.org

Charles Gale
Harvard Graduate School of Education; chagale@gmail.com

Katrina Kosec
International Food Policy Research Institute; k.kosec@cgiar.org

This study benefited at various stages from experts at the World Bank, the International Food Policy Research Institute (IFPRI, the Tanzania Social Action Fund (TASAF and elsewhere. At TASAF, the evaluation has been supported by the Executive Director Ladislaus Mwamanga, as well as the former Executive Director Servacius Likwelile. Amadeus Kamagenge led TASAF input to the evaluation, and his entire team has contributed with substantive and logistical support. At the World Bank, Samantha de Silva, Myrtle Diachok, and Ida Manjolo provided crucial support. We received financial support from the CGIAR Research Program on Policies, Institutions, and Markets led by IFPRI, the International Initiative for Impact Evaluation (3ie, the Strategic Impact Evaluation Fund (SIEF, the Trust Fund for Environmentally and Socially Sustainable Development (TFESSD, and the Bill Melinda Gates Foundation. Each author contributed equally to the paper. This work benefited from helpful comments from Sarah Baird, Felipe Barrera-Osorio, Shelby Carvalho, Emmerich Davies, Daniel Gilligan, Lawrence Katz, Sikandra Kurdi, Paul Niehaus, Julie Schaffner, Eric Taylor, Rebecca Thornton, and two anonymous referees. Amina Mendez Acosta and Emma Cameron provided valuable research assistance.

The data from the Tanzania conditional cash transfer pilot are publicly available at **https://microdata.worldbank.org/index.php/ catalog/2669**. The data reported in Table 8 (Aspects of Heterogeneity Analyzed in 101 Earlier Studies of Cash Transfers on Educational Outcomes) are available at **https://www.cgdev.org/sites/default/files/2022-07/Evans-heterogeneity-cash-transfer-data-revised-format.xlsx**

David K. Evans, Charles Gale, and Katrina Kosec. 2022. "The Educational Impacts of Cash Transfers in Tanzania." CGD Working Paper 563. Washington, DC: Center for Global Development. **https://www.cgdev.org/publication/educational-impacts-cash-transfers-tanzania**

CENTER FOR GLOBAL DEVELOPMENT
2055 L Street, NW Fifth Floor
Washington, DC 20036

1 Abbey Gardens
Great College Street
London
SW1P 3SE

www.cgdev.org

1 Introduction

With the attention of international institutions turning to the quality of global education (World Bank, 2018, 2019a), one might imagine that access to school is universal. However, the completion rate for primary school in low-income countries is just 65 percent: a large minority of children still fail to complete primary school or to reap the gains of education.[1] How can policymakers and donors help the last third of children in the world's poorest countries complete their primary education? Cash transfers, which are usually targeted to low-income households, have been used around the world with the goal of improving educational outcomes, and many studies have verified that they increase educational access for the average beneficiary child, in both the short and long runs (Baird et al., 2014; Molina Millán et al., 2019). But as countries seek to expand educational access to the remaining out-of-school children, evidence on their effectiveness on specific educational outcomes and at reaching children who face particularly high levels of poverty and multiple vulnerabilities grows more important.

In this paper, we test the impact of cash transfers on schooling outcomes for low-income households in the low-income country context of Tanzania.[2] We also test the efficacy of cash transfers for children facing varied challenges beyond poverty alone. These include gender-specific challenges, challenges facing children from the very poorest households, and challenges facing children with poor initial school performance. We then situate these impacts within the context of one hundred previous evaluations of how cash transfers affect education, drawing on two existing reviews and our own descriptive analysis of heterogeneous treatment effects from the underlying studies. To test the impact of the program, we use a randomized controlled trial of conditional cash transfers targeted by and delivered through community committees in Tanzania, a country with a 68 percent primary school completion rate, just above the average for low-income countries.[3] In order to receive the transfers, households needed to ensure that children aged 7–15 were enrolled in school and attended at least 80 percent of the time. They also had to meet required numbers of health clinic visits (Evans et al., 2019a).

The cash transfers had the potential to affect children's education in several ways. First, the conditions themselves may have boosted attendance as households may have sought to satisfy the conditions to qualify for the transfers. Second, the conditions may have served as a nudge to households to focus on education, even beyond satisfying education requirements. (There is some evidence that the same program boosted health-seeking behavior above the requirement of the program (Evans et al., 2017).) Third, the additional resources to households may have made education more feasible by relaxing other constraints or by complementing other household resources for education.

[1] Statistics for primary school completion in this introduction are from World Bank (2019b) and reflect the latest available data.

[2] Tanzania was classified as a low-income country until mid-2020, at which point the World Bank reclassified it as lower-middle income (Battaile, 2020).

[3] The primary completion rate of 68 percent is for 2020.

With a limited initial budget, the Government of Tanzania worked with researchers to pilot conditional cash transfers in 40 randomly assigned villages, with another 40 villages serving as a comparison group. We compare beneficiaries in treatment villages with non-beneficiaries in comparison villages after 1.75 years of transfers (our midline) and again after 2.75 years of transfers (our endline). We correct our results for potential false positives due to multiple hypothesis testing.

Our study focused on children currently aged 6–21; in our sample, this encompasses the full range of ages of children attending either primary or secondary school.[4] We find that cash transfers boosted the likelihood that children between the ages of 6 and 21 had ever attended school by between four and five percentage points at both midline and endline. Suggestive evidence indicates that this is driven by higher rates of age-appropriate enrollment for the youngest children. (In other words, the youngest children are less likely to enroll later than legal guidelines recommend.) Overall, we observe no impacts on children's absenteeism or primary completion rates, though we do find that students with stronger initial educational performance experience increases in primary completion rates. While we observe no statistically significant impacts on completion of the first year of secondary school, we do observe negative point estimates for that outcome, which may be an indication that households in the program focused their efforts on keeping primary school aged children – those for whom the program conditions were binding – in school.

We test for heterogeneity in two ways. First, we test for heterogeneity across three baseline parameters with a strong theoretical basis: gender, poverty, and educational performance. Households may respond differently to cash transfers by gender given that returns to education may differ by gender (Psacharopoulos and Patrinos, 2018) and given that many studies establish differential spending across genders, albeit not always favoring one or the other (Aslam and Kingdon, 2008; Dizon-Ross and Jayachandran, Dizon-Ross and Jayachandran; Masterson, 2012; Wongmonta and Glewwe, 2017). Other studies model those differences based on differences in costs and in return experienced by parents (Alderman and King, 1998; Pasqua, 2005). Poverty also has a basis for likely heterogeneous impacts of transfers. Previous research has established a positive elasticity of household investments in education with respect to income, but with different elasticities in different contexts (Acerenza and Gandelman, 2019; Ogundari and Abdulai, 2014). As such, households with higher baseline wealth may respond differently to similar absolute increases in transfers. Finally, initial educational performance and other indicators of perceptions of children's ability can determine how households invest in children's education, as theorized by Becker and Tomes (1976) and demonstrated by Dizon-Ross (2019).

Second, beyond these well-motivated characteristics, we use machine learning (specifically, a causal forest algorithm) to test for heterogeneity of effect sizes across a larger set of child and

[4] As Figure A1 illustrates, children identified as attending primary school are 6-16 years old, while children identified as being in secondary school are 14-21 years old, and the official age ranges for primary school and secondary school are 7-13 and 14-19, respectively.

household characteristics. Doing so helps us avoid allowing our priors to determine the potential sources of heterogeneity. We consider both which characteristics are most important in accounting for heterogeneities (looking at variable importance across different outcomes) as well as which combinations of characteristics are most important (we create two-dimensional heat maps allowing us to see which combinations of two characteristics account for the largest conditional average treatment effects).

We observe similar impacts for girls and for boys for whether they ever attended school, for school attendance, and for primary completion. We observe a significant, negative impact on completion of the first year of secondary school only for boys. We also observe that the benefits on ever having enrolled are concentrated among the less-poor children in our overall poor sample, and among those with weaker educational outcomes at baseline.

This research contributes to two principal literatures. The first is on the effectiveness of cash transfers in achieving human capital objectives. A recent review identifying 20 studies measuring the impact of cash transfers on school attendance found that nearly two-thirds found significant positive impacts (Bastagli et al., 2016; Hagen-Zanker et al., 2016). A meta-analysis likewise found positive impacts for both conditional and unconditional transfer programs (Baird et al., 2014). However, few of those studies separately identified impacts for the poorest children or for those who faced other vulnerabilities. From a sample of more than one hundred cash transfer evaluations with educational outcomes drawn from Bastagli et al. (2016) and Baird et al. (2014), we find that while most (71 percent) report outcomes separately by gender, only 29 percent differentiate by poverty level within cash transfer recipients, and only 7 percent by baseline student performance. Our study shows that conditions may not by themselves overcome all constraints for the most vulnerable children, and how the effectiveness of transfers varies across children with different vulnerabilities.

The second relevant literature explores ways to ensure educational access for the most vulnerable children, often facing a variety of vulnerabilities. Recent research has examined a range of interventions to reach children who remain out of school. Some of these focus on reaching children in the poorest communities (Fazzio et al., 2020) or ensuring that the most vulnerable girls (Psaki et al., 2022; Sabates et al., 2020) and the most vulnerable boys (UNESCO, 2022; Welmond and Gregory, 2022) remain in school, while others focus on boosting educational outcomes for orphans (Cho et al., 2017; Hallfors et al., 2011; Thomas et al., 2020) or for lower performing children (Banerjee et al., 2017; Duflo et al., 2020). Our study examines the efficacy of a commonly used policy instrument in boosting outcomes for children with a range of vulnerabilities.

These findings suggest that to achieve truly universal school enrollment, policymakers will likely need to draw on an array of interventions to get the most vulnerable children into school and to help them stay there. Further, the relative importance of different vulnerabilities—gender, relative poverty, and academic performance among them—will vary across contexts.

While our results provide insights on the impact of cash transfers across educational outcomes and different groups, they come with limitations. We observe outcomes 2.75 years after the initiation of transfers (the timing of our endline survey), but the ultimate objectives of school completion and improved life outcomes come much later. A handful of studies identify the long-term human capital impacts of cash transfer programs (Molina Millán et al., 2019; Barrera-Osorio et al., 2019; Molina Millán et al., 2020; Araujo et al., 2019), and more work is needed on identifying those impacts, particularly for the most vulnerable children and youth. In addition, while receipt of cash transfers was randomly assigned, vulnerabilities are not. As such, our heterogeneous treatment effects have a non-causal interpretation; that is, any given dimension of heterogeneity may capture other factors correlated with that dimension.

The rest of this paper proceeds as follows. Section 2 characterizes the context and the intervention. Section 3 describes the data, and Section 4 discusses the identification strategy, balance tests, and analyses of attrition. Section 5 presents the results. Section 6 discusses the implications of our findings in the context of two reviews of over one hundred papers. Section 7 concludes.

2 Background

2.1 Context

Tanzania is a country with high rates of poverty and other vulnerability. As of 2018, more than one in four people were classified as poor. Ten years earlier, close to the start of the program we evaluate, that number was more than one in three (Belghith et al., 2019). At the same time, many children face vulnerabilities in addition to poverty. In many countries in Sub-Saharan Africa, girls continue to complete fewer years of education than boys (Evans et al., 2021), although recent evidence from Kenya, Tanzania, and Uganda suggests that girls outperform boys (on average) in both numeracy and literacy tests (Buhl-Wiggers et al., 2021). As such, vulnerabilities by gender may not be consistent across educational outcomes. Overall, learning outcomes are low in Tanzania. One national measure of literacy suggests that only 10 percent of children in the third year of primary can read at a second grade level, and only 39 percent of fifth graders can read at a second grade level. For numeracy, only 23 percent of third graders (and 54 percent of fifth graders) could perform at a second grade level (Uwezo, 2019). Combinations of these vulnerabilities can result in multiple exclusions and further decreases in access to educational opportunities (Moodley and Graham, 2015; World Bank, 2018).

Cash transfers may interact distinctly across these different vulnerabilities. Transfers most obviously alleviate a poverty constraint. However, in the face of relaxed constraints, parents may still make choices about which children's education to invest in more heavily. How cash transfers affect educational outcomes for different children and youth will depend on the interactions between relaxing the poverty constraint (to a degree), imposing conditions, nudging households towards greater attention to education, and other characteristics of the child and the household.

In this context, we collaborated with the government of Tanzania to evaluate the impact of cash transfers. The Tanzania Community-Based Conditional Cash Transfer (CB-CCT) pilot program was conceived in 2007 in discussions between the Tanzania Social Action Fund (TASAF) and World Bank officials. TASAF is a government agency tasked with promoting economic and social development,[5] established in 2000. In its early stages, the program provided funding and support to communities who applied for support in implementing locally-managed infrastructure projects, such as rehabilitating schools and health clinics. Despite receiving extensive donor funding, according to Hickey et al. (2019), the emphasis on community-based programming reflected stated government goals of self-reliance and community participation in markets. The CB-CCT pilot was one of several projects the government wanted to test for possible scale-up,[6] and the government enlisted the help of World Bank researchers to evalute it.[7]

2.2 Selection of Villages

The pilot took place in 80 eligible villages in three districts in Tanzania—Chamwino, Bagamoyo and Kibaha. Two of these districts are relatively close to the largest city, Dar es Salaam, while the third is close to the capital, Dodoma.[8] TASAF prioritized villages that had successfully managed projects in the past, and thus had training on and experience with procurement, budgeting, and contracting.[9] The 80 villages were randomized into treatment and control groups of 40 villages each, stratified on village size and district. The CB-CCT program was implemented for a period of three years (2010–2012) in treatment communities and then extended to all 80 communities shortly after our last round of data collection (in late 2012). After village leaders and citizens were notified of the results of randomizing villages into treatment and control, control villages were told they would begin receiving the program in late 2012, while treatment villages were told that they would continue receiving the program indefinitely (and past 2012).

[5]Though nominally pro-poor, Baird et al. (2013) find regressive elements in the process by which villages apply to the program, in that communities with knowledge of program availability and requirements may be slightly better off than the poorest of the poor. In this way the program should be thought of as "moderately pro-poor" rather than explicitly.

[6]In 2012 the government began the Productive Social Safety Net (PSSN), a nationwide CCT that aimed to eventually reach five million people (Hickey et al., 2019) that was informed by preliminary findings from the CB-CCT pilot.

[7]In separate research, the project was found to have improved health outcomes (Evans et al., 2019a) as well as social capital, in the form of feelings of trust toward government (Evans et al., 2019b) and other citizens (Evans and Kosec, 2020).

[8]At baseline, villages ranged from 64 to 10,078 households. The average size was 980 households, and the median size was 560 households.

[9]All 80 villages participating in the CB-CCT pilot were further slated to *not* receive other TASAF or special government or donor funding during the duration of the pilot, to avoid confounding effects.

2.3 Selection of Households

In each village, prior to randomization into treatment and control, citizens elected members of a community management committee (CMC). Their job was to identify beneficiaries, administer and oversee the program, and enforce the program conditions. Once elected, CMC members were trained to undertake a survey where they collected data on the housing conditions and access to food for approximately the poorest 50 percent of households in each community. TASAF fed this information into a proxy means test and then provided a ranked list of households to the CMCs; funding availability established a line separating a set of proposed beneficiaries from other households. CMCs generally certified the list; in some cases, they petitioned TASAF for small modifications.

Village-level randomization then occurred. For villages randomized into treatment, targeted households and their members became program beneficiaries, while targeted households and their members in control villages became "would be" beneficiaries. Household poverty status was verified on a sample basis by TASAF. On average, the number of beneficiary households represented 23 percent of the total households in each village. Due to a combination of household refusal and last-minute changes in community decisions, roughly 9 percent of households in treatment villages did not receive treatment.[10] Data from the midline survey showed that 58 percent of households reported a CMC member was a neighbor, and 23 percent reported that a CMC member was a blood relative.[11]

2.4 Intervention

TASAF began delivering cash transfers in January 2010. Payments to households were made every 2 months for the duration of the pilot (2 years and 9 months, or from January 2010 through September 2012), as long as households remained eligible. The amount of the transfer was determined based on the food poverty line, and was—when the program initially rolled out—3,600 Tanzanian shillings per month for each child up to 15 years of age, and 7,200 TSh per month for elderly persons aged 60 and over. (All amounts in this section are reported in constant 2009 Tanzanian shillings.) To our understanding, TASAF confirmed the ages of children in beneficiary households (would-be beneficiary households for the case of control communities) via data collected by community management committees. We did not hear of any concerns or registered complaints that parents were able to misreport child ages. At midline, the median most recent transfer payment was reported to be 18,800 TSh.[12] Transfers amounts were adjusted over time, and the median most recent transfer payment at endline was 21,610 TSh per

[10] These individuals remain in our sample.

[11] The household survey data are available through the World Bank microdata catalogue, at https://microdata.worldbank.org/.

[12] We use consumer prices index averages for Tanzania for 2009, 2011, and 2012 from World Bank (2019b) to convert all transfer amounts into constant, 2009 Tanzanian Shillings (TSH).

household, which would equal roughly 130,000 Tsh over the course of a year, given six annual transfer payments. At both midline and at endline, the median most recent transfer payment households had received was equal to roughly 13 percent of total household expenditures over the same (two month) period.

Total annual household education expenditures increased over the evaluation period.[13] (We calculate the expenditures for households with at least one 6 to 12 year old at baseline). These households were spending about 44,000 TSh in annual education expenditures at baseline, on average. This amount increased to 78,000 TSh by endline in control villages, while it increased to 63,000 TSh in treatment villages. For households with at least one 13 to 18 year old at baseline, annual education expenditures increased from 48,000 to 81,000 TSh between baseline and endline in control villages, and from 37,000 to 72,000 TSh in treatment villages over the same period. Thus, the transfers were more than enough to cover all annual educational expenditures. (Transfers were not solely targeted towards educational outcomes.)

While payments were made at the household level, conditions applied at the individual level. The conditions were that children age 0-5 would visit health clinics at least 6 times per year, children age 7-15 be enrolled in school with 80 percent attendance, and elderly persons visit clinics at least 1 time per year.

The academic calendar follows two terms, the first from January to June and the second from July to November. Primary education, which is compulsory, lasts 7 years. School starting age—which in our study context is age 7 for over 40 percent of children, and age 8 for over 20 percent (Figure A2; the official starting age is 7)—does not differ drastically across the different vulnerabilites we consider in this paper.

The CMC oversaw monitoring, which involved collecting forms from health clinics and schools and maintaining records, delivering warnings to households that were found to be out of compliance, and making regular visits to beneficiary households. Monitoring was conducted every four months and records were submitted to TASAF and entered into a centralized database, where a final payment list was generated. At baseline, 99.7 percent of 7–15 year olds attending school attended government or community schools; 0.3 percent attended either a religious or other private school. This facilitated government collection of attendance records.

According to program rules, if households were found to be out of compliance (e.g., due to any of the children aged 7–15 in the household registered as beneficiaries not attending school 80% of the time), a first warning was issued, and if at the next monitoring period (8 months after the first payment), beneficiaries still failed to comply with all conditions, payments were reduced by 25 percent and a second warning was sent. After two warnings were issued, beneficiaries that failed to fully comply were suspended indefinitely, but could return to the program after review and approval by the CMC and TASAF.

[13]This was measured as the total amount spent on an individual child's education by household members over the previous 12 months, and included school fees, books, materials, uniforms, transport, extra tuition or school contributions.

Our qualitative work via key informant interviews with village leaders, CMC members, head teachers, and healthcare workers indicated that one perceived key to high compliance was that the individuals doing the monitoring were actively engaged with community members—taking advantage of the local nature of monitoring. For example, village leaders and CMC members described visiting homes to urge parents to send children to school to avoid loss of transfers, while healthcare workers described encouraging parents to buy health insurance immediately after they got they transfer payments (when they "still felt rich") (Evans et al., 2017). While we do not have precise data on the strictness of enforcement, few households (2–3 percent) reported receiving payments that were smaller than usual due to program non-compliance during the midline and endline surveys. It is possible that enforcement was stricter for some conditions (school enrollment) than others (school attendance), although we have no evidence that this was the case. Households also reported high satisfaction with their CMCs. Roughly 93 percent of households reported receiving their transfer from a community office during the midline.

3 Data

In this section we describe the data collection process. The impact evaluation comprised three waves of household-level data collection. The baseline was carried out during January–May 2009, prior to the January 2010 start of transfer payments. The midline was carried out from July–September 2011 (approximately 2.5 years after baseline, and 1.75 years after treatment began), and the endline was carried out during August-October 2012 (approximately 3.5 years after baseline, and 2.75 years after treatment began). The endline concluded prior to the extension of transfers to control villages. Table 1 presents the chronology of both the program and the evaluation.

The baseline survey included 1,764 households including 6,918 individuals. The number of households to be interviewed per community was set at 25, determined by power calculations and known effect sizes from other CCT studies. Over 90 percent of respondents to the household survey are either the household head or the spouse of the household head. Detailed individual and household-level data were collected on consumption, health-seeking behavior, education, and other factors. During the endline survey, enumerators also collected data on schools and health clinics in treatment and control villages. For the present analysis, we work with an analytic sample comprised of 1,064 households and 2,424 individual-level observations aged 6-21.

Members of the enumeration team were not identified or affiliated with representatives from TASAF or the CMCs, and all interviews were conducted without TASAF representatives present. In addition, steps were taken during questionnaire construction to limit respondent bias. For example consumption items were grouped into lists and not singled out or highlighted, such as in the case of temptation goods (i.e., cigarettes or alcohol).

3.1 Outcomes

The main individual-level education outcomes analyzed in this study are related to school attendance—specifically, whether children have ever attended and rates of absenteeism (3 outcomes)—and school progression and completion (2 outcomes). Unfortunately, we lack data on school performance (e.g., test scores).

Variables related to school attendance include: ever attended, which is an indicator for having ever attended school;[14] missed last week, which is an indicator for the child being reported by the main respondent (typically a parent) as having missed school at least once in the last schooling week for a reason other than a public holiday, school closure, or teacher absence (it also takes a value of 1 if the child is not enrolled in school) (we loosely refer to this as parentally-reported attendance); and school attendance, which is an indicator variable only present in the endline survey for the child being recorded as attending more than 80% of open school days in the May 2012 school register (administrative records kept by the head teacher).

The cash transfer was conditioned on households keeping their children between the ages of 7 and 15 enrolled and attending school at least 80 percent of the time. Though there were multiple, widely known mechanisms for auditing the attendance of children in the community, it is possible that households misreported attendance or other education outcomes. Most obviously, parental reports could be influenced by social desirability bias, thus providing a noisy signal of true absences. More problematically, this social desirability bias may be greater specifically in treatment villages, as parents are aware that attending school 80 percent of the time is a condition of the program, and they may have perceived enumerators to be connected to TASAF or the government. This could bias upward estimates of the effect of the CCT in reducing absenteeism (i.e., it may appear that the CCT reduces absenteeism by students more than it actually does). Further, a single week cannot capture broader patterns of attendance over a longer period—and may be a noisy measure of a child's typical attendance rate if, for example, school was closed for several days in the past week (thus reducing the number of "open" days during which attendance decisions could be recorded). This motivated us to use administrative records of attendance as a second source.[15]

During endline data collection, the research team sent enumerators to all primary schools (though not secondary schools) in treatment and comparison villages that were attended by children in the village during the midline survey. They collected administrative data on recorded child presence in school during May 2012 (i.e., the last full month which pre-dated our endline survey and for which school was generally in session across Tanzania). The school year begins in January and has two terms (January to June and July to December; as a result, May would be in the latter part of the first term (Shupavu, 2020). We compare analysis using the parentally-

[14] Figure A2 shows that the modal school start age is 7, with both 8 and 9 being more common than 6.

[15] Direct comparisons of parental attendance reports and school records in other studies suggest that these can yield very different estimates, hence the value in including both in our main analysis (Evans and Mendez Acosta, 2021).

reported measure of absenteeism with this administrative attendance measure, acknowledging the potential limitations of each.

Progression and completion variables include primary completed, which is an indicator for the child being reported as having completed primary school; and first grade of secondary completed, which is an indicator for the child being reported as having completed at least one grade of secondary school.

Three of our outcomes—whether children ever attended school, whether children completed primary, and whether children completed at least one grade of secondary—are represented with indicator variables that, upon becoming a 1, can never turn back to being a 0. It is also the case that if any of these indicators is 0 in a later period, then it also must be 0 in an earlier period. This logic allows us to replace missing values with known, non-missing values (1 or 0) for some children who may have been missing from the sample in one or more periods—and accordingly reduces sample attrition for these outcomes. Appendix A.1 contains more details.

3.2 Selection of ages for analysis

Our analytic strategy requires identifying which child–year observations to include when estimating treatment effects. A useful first step is considering the official ages for primary school and secondary school in Tanzania and the actual age range of students in primary school and in secondary school in our baseline (pre-treatment) dataset. These age ranges are displayed in Figure A1. The official ages for primary school are 7–13 and the official ages of secondary school are 14–19 (assuming age-appropriate start and no grade repetition). Meanwhile, the age ranges of children identified as being in primary school and in secondary school in our sample at baseline are 6-16 and 14-21, respectively.

When our outcome is an indicator for having ever attended school or either of our measures of attendance (i.e., absenteeism)—either the parental or the administrative report—our main estimation sample includes children who were currently aged 6–21 (inclusive) at the time of each survey round.[16] This reflects the fact that we anticipate meaningful variation in these outcomes for all children, whether they are in primary school or secondary school, and this age range captures all primary and secondary school students in our sample. We note that this age range encompasses children subject to the program conditions (which affected 7–15 year olds) as well as children who may benefit indirectly from a higher household income. We also report, however, similar results for the subset of children aged 7–15, for whom program conditions applied (see Table A1).[17]

The CB-CCT pilot did not have an explicit condition on children completing primary school,

[16]Note that our focus on current ages, in contrast to fixing a cohort of children based on baseline ages, implies that a given child may age out of our estimation sample, or age into our estimation sample.

[17]Our administrative attendance data for the month of May 2012 (just before the endline survey) were collected only from primary schools, and thus contain a few observations of children over age 15.

but this was a goal of the program. Given that secondary school students in our sample range in age from 14–21 at baseline (and 13 is indeed below the age that secondary school officially begins), we chose those children who were currently aged 14–21 (inclusive) at the time of each survey round when studying outcomes related to having completed primary school and having attended secondary school. Youth aged 14–15 years old would have been directly incentivized to stay in school, while 16–21 year olds might indirectly benefit from having greater resources in their household from the transfers. Of course, primary completion may occur before age 14 for some students, and we do observe a smaller sample of secondary school students under age 14 in the midline and endline surveys (albeit not in the baseline). For that reason, we include a robustness check that includes younger students.

4 Estimation Strategy

Given that assignment to treatment was random, we can estimate the causal intent-to-treat effect of the CCT. We estimate midline and endline effects for the CCT treatment using the following pooled specification:

$$Y_{ihvt} = \alpha_0 + \beta_1 (T_v * M_t) + \beta_2 (T_v * E_t) + \delta_1 M_t + \delta_2 E_t + \gamma_1 M_t \times \mathbf{X_{ihv}} + \gamma_2 E_t \times \mathbf{X_{ihv}} + \eta_i + \epsilon_{ihvt} \quad (1)$$

where Y_{ihvt} is the outcome for individual i in household h and village v at time t, M and E are survey phase dummies for the midline and endline respectively, T_v=1 if the individual lives in a treatment village and 0 otherwise, η_i are individual fixed effects, X_{ihv} is a vector of the baseline values of covariates that are imbalanced at baseline, and e_{ihvt} is an idiosyncratic individual-level error term. We recover causal estimates of the effects of treatment at midline and at endline through the coefficients β_1 and β_2, respectively. All regressions include standard errors clustered at the village level, the unit of randomization. We examine impacts on all five of our outcomes overall and by each of our three theoretically-motivated dimensions of vulnerability.

4.1 Outcome of the randomization

While villages were randomized into treatment and control, it is possible that some characteristics of treatment villages, or of households and individuals within them, are significantly different than their control village counterparts. If imbalances were present, one might worry that they—rather than treatment—explain differences in outcomes. We address this concern in two ways. First, we show that randomization generally led to balance across treatment and control villages. Second, as shown in Equation (1) in Section 4, we use individual fixed effects in addition to interactions between the baseline values of imbalanced covariates and year dummies, to account for baseline imbalances.

In Table 2, we examine differences in baseline means between treatment and control groups for an array of individual child, household, and village characteristics. (For child characteristics, we consider 6–21 year olds when we test for balance across demographic characteristics and those outcomes analyzed for 6–21 year olds–such as whether they ever attended school. Wwe consider 14–21 year olds for outcomes analyzed for 14–21 year olds–such as whether a child completed primary school.) In total, we examine 22 characteristics; we find baseline imbalances at the 0.10 level of significance or higher for only two characteristics: completed the first grade of secondary school (where the treatment group was 7 percentage points less likely to have completed it) and number of household members 60 and over (where the treatment group had 0.16 fewer such members). We deal with these imbalances by always estimating models with individual fixed effects (these absorb not only these imbalances, but also the effects of all other time-invariant unobservables) as well as controlling for each child's baseline value of these two imbalanced covariates interacted with year dummies, thus allowing children with different baseline values to be on different, potentially non-linear, trends for our outcomes over time.

Despite these imbalances, we find balance for a large array of potentially important characteristics and outcomes at baseline. Examining household-level characteristics, we find no imbalances on whether or not the head has education, whether or not the household has below-median consumption, or whether the head believes school quality is good or excellent. Considering village characteristics, we further see that treatment and control villages do not exhibit any statistically significant differences with regard to size (number of households), connectedness to higher levels of government (whether the village houses the ward, or whether the village executive officer (VEO) lives in the village), frequency of village meetings (number held in the last year—where local leaders are supposed to hold four), or poverty level (specifically, the village's poverty rank in TASAF's database, or the village's poverty score in TASAF's database).

4.2 Attrition

Between baseline and midline, 0.2 percent of households attrited from the sample, and between baseline and endline, 0.8 percent of households attrited. Table 3, columns 1–4 consider correlates of a household remaining in the sample at midline (columns 1–2) and at endline (columns 3–4). We estimate two specifications for each follow-up round: one with only district fixed effects and an indicator for living in a treatment village, and another with district fixed effects, an indicator for treatment, baseline household-level characteristics (including head gender, head age and age squared, and dummies for the head having no education and for the household having below-median household consumption), and interactions of treatment with these baseline characteristics.

As columns 1–4 show for the case of household-level attrition, in no cases is treatment, or its interaction with any baseline household-level covariates, a statistically significant predictor of a household remaining in the sample. While low consumption households are significantly more likely to remain in the sample, we cannot reject that consumption has the same effect on remaining in the sample across treatment and control villages. For each of the regressions with interactions with treatment, we further compute the F-statistic for the joint significance of treatment and all of these interaction terms. At both midline (column 2) and endline (column 4), we can reject that treatment and its interactions with covariates jointly predict a household remaining in the sample. Overall, we conclude that treatment does not affect which households remain in our sample.

We additionally conduct a similar, individual-level attrition analysis for children aged 6–21 at baseline (the age range we study) in columns 5–8. For this analysis, in addition to interacting treatment with household-level variables, we also interact it with the child's own gender, age, and age squared. We see that the indicator for treatment is not a statistically significant predictor of an individual remaining in the sample. Further, among the various baseline characteristics (four at the household level, two at the individual level), only for one—age of the household head—is the interaction between treatment and that characteristic itself statistically significant. We also find that treatment and its interactions with covariates are not jointly significant predictors of an individual remaining in the sample, either at midline (column 6) or at endline (column 8)—just as for the case of household-level attrition.

4.3 Adjustments for multiple hypothesis testing

Because this program tests five education outcomes and considers a number of heterogeneous treatment effects which were identified after the program was underway, we correct our findings for potential false-positive results due to testing multiple hypotheses. The Romano-Wolf method controls the familywise error rate (FWER) and allows for dependence among p-values through the use of bootstrap resampling (Clarke et al., 2020). From this procedure, we obtain Romano-Wolf adjusted p-values, which we indicate in brackets beneath the clustered standard errors in our regression tables or—in the case of p-values for tests of differences between coefficients—as rows (indicated with RW) at the bottom of our regression tables.[18]

Correction for multiple hypothesis testing is sensitive to how hypotheses are grouped, underscoring the importance of clarifying and motivating this grouping. Our paper groups hypotheses according to the specific dimension of heterogeneity we are examining, and thus by table, combining those related to coefficients themselves (at midline and at endline) as well as tests for differences between different vulnerable groups.

[18]Controlling the FWER is most appropriate for cases where the cost of a false rejection of the null hypothesis has strong policy implications; an alternative approach is to control the false discovery rate, which limits the expected proportion of false positives (Anderson, 2008).

5 Results

In this section, we present the impacts of Tanzania's CCT program on a variety of education-related outcomes. First, we consider outcomes related to school attendance—specifically, whether children aged 6–21 have ever attended school and rates of absenteeism. Second, we examine impacts on school progression and completion for children aged 14–21. Third, we consider how impacts vary across theoretically motivated individual- and household-level markers of vulnerability (i.e., child gender, household consumption, and baseline exam performance) and then also across a wide range of other characteristics using machine learning.

5.1 Attendance

Participating in the CCT program boosted the likelihood that children aged 6–21 ever enrolled in school (Table 4). At midline, treatment led to a 4 percentage point increase in the likelihood of a child aged 6–21 having ever attended school (significant at the 0.05 level using clustered standard errors, and with a Romano-Wolf adjusted p-value of 0.01), with the same effect size and level of statistical significance at endline.[19] The baseline likelihood of having ever attended school was 81 percent, so this reflects about a 5 percent increase in the likelihood of any child in this age range having attended school, relative to the baseline mean.

Despite a greater likelihood of having ever attended school, we find little evidence that the CCT program reduced student absenteeism. Our regressions do not pick up a statistically significant treatment effect on either of our two absenteeism measures (Table 4). For our first measure, we consider a measure of absenteeism using parental reports of whether the child (aged 6–21) missed school in the last schooling week (column 2). For our second measure, we use an indicator for a child enrolled in primary school being reported as having attended 80 percent or more of open school days during the month of May 2012 (column 3; this is a variable similar, but not identical, to the program condition of 80 percent attendance on average during the school year). These estimates do not condition on school enrollment, and thus students who are not enrolled in school receive a 0 for both indicators. Further, the estimates using administrative reports rely on only endline data and thus use district rather than individual fixed effects. Both measures of absenteeism suggest null effects of the CCT.

We obtain similar results for attendance-related outcomes when estimating our main specification without controlling for the baseline values of our two imbalanced (at baseline) covariates interacted with year dummies, as shown in Table A2. We obtain a larger (5 percentage point) endline treatment effect on an indicator for having ever attended school (column 1), but obtain similarly null results for our two absenteeism indicators. Figure A3 plots the density of missed days in May 2012 (i.e., just before the endline survey) by treatment group. While we observe a

[19] We find a similar effect when considering a smaller subset of children aged 7–15, which is the subset of children for whom attendance conditions applied (see Table A1).

greater density at zero days absent for the treatment compared to the control group—consistent with a reduction in absenteeism due to the program, there is a slightly wider right tail for the treatment group.

Our effect sizes may vary with baseline age, which gives some insight into how to properly interpret the results. We show this in Figure A4. There, we see that the CCT did not increase the likelihood of having ever attended school across the full age distribution; rather, our statistically significant effect sizes for this outcome are driven by increases concentrated among 6 and 7 year-olds. Given the distribution of school start ages, shown in Figure A2, this suggests that the program may have incentivized earlier enrollment rather than expanding enrollment of all children. Since all of these children were of school age, this suggests that the program may have increased age-appropriate enrollment of younger children.

We also see that our null results for whether or not the child's parents reported them as having missed one or more days of school in the last week hides a pattern whereby there are declines in absenteeism in very young children and possibly some older children—though we see nearly significant *increases* in absenteeism among children at the age of transition into secondary school (those with baseline ages of 13, who would be roughly 15 at midline and 16 at endline). This is perhaps unsurprising given that the conditions of the programs applied only to children 15 and under, and not those older than 15. (But given that we lack the power to estimate age-specific absenteeism with great precision, we don't make too much of it.) However, for our administrative measure of absenteeism, we see little effect of the CCT program for any age group.

5.2 School progression and completion

We next consider the impacts of the CCT program on school progression and completion among children aged 14–21 in columns 4–5 of Table 4. (Children younger than this age group would not be expected to have completed primary school, given the distribution of ages we observe in Figure A1). While primary school is nominally free and compulsory, the transition to secondary school and beyond comes with higher parental monetary contributions, longer travel times to school, and other opportunity costs (as children age and are capable of carrying out more tasks). We consider two outcomes: an indicator for the child having completed primary school (column 4) and an indicator for the child having completed at least one grade of secondary school (column 5).

We find no clear evidence that the CCT program affected either the completion of primary school by children aged 14–21 or their likelihood of completing at least one grade of secondary school. On average, while the CCT program boosted school attendance, it did not appear to affect progression or completion overall. The point estimates for primary completion are close to zero. The point estimate for completion of the first year of secondary at endline is -0.09 but statistically insignificant, suggesting the potential of an adverse impact but without the statistical power to confirm that.

As a robustness check, we consider the impacts of transfers on our progression and completion outcomes when estimating our main specification *without* controlling for the baseline values of our two imbalanced (at baseline) covariates interacted with year dummies; these estimates are shown in Table A2, columns 4–5. We obtain null effects of treatment that are similar in magnitude to our main estimates of Table 4.

As an additional robustness check for our progression and completion outcomes, we widen the age group considered from those currently aged 14–21 to those currently aged 12–21. In doing so, we expand the number of years that a child in early adolescence appears in our panel (e.g., given we focus on a child's current age, then a child who is 14 at endline would appear in our dataset only at endline and not in earlier rounds when we fix our age range as 14–21, and thus they would not influence our treatment effect estimates in our individual fixed effects model, whereas that child would appear in two or possibly even all three rounds if we fix our age range as 12–21). As for Table 4, We do not identify any benefits of the CCT program for this age group (see Table A3). In fact, we observe a statistically significant, negative effect of the CCT at endline (though not at midline) on having completed the first grade of secondary school—the same 9 percentage point decline we observed earlier, but this time with statistical significance. This may suggest that households are shifting resources away from secondary school-aged children (who are not incentivized by the program to attend school) and instead toward young, primary-school-aged children (on whose attendance transfers depend).

5.3 Heterogeneous effects by common markers of vulnerability

Our overall estimates suggest increases in school attendance, but minimal effects on progression or completion overall. In this section, we examine how the results vary across children who do and do not exhibit common markers of vulnerability. Specifically, we consider three theoretically-motivated markers: child gender, household poverty (i.e., whether or not the household has below-median consumption), and an indicator of baseline school performance (i.e., whether the individual had passed a Standard IV—i.e., 4th grade—exam at baseline or not; those who did not even sit for this exam are coded as not having passed it). We see in Table A4 that these three indicators of vulnerability are all positively but quite weakly correlated with one another, with correlation coefficients always below 0.1.

5.3.1 Gender

In the case of gender, we find no systematic evidence of statistically significantly different impacts on boys compared to girls. Table 5 shows that at both midline and endline, the point estimate of the effect of treatment on girls' likelihood of having ever attended school is identical to and statistically indistinguishable from that for boys (whether using conventional or Romano-Wolf adjusted p-values), despite the effect being statistically significant at conventional levels at midline and at endline for girls but not boys (though at endline, the Romano-Wolf adjusted p-value of the effect for girls is 0.13—just below conventional levels).[20] Further, our results for absenteeism and for primary completion are statistically insignificant for both girls and boys. For progression to secondary, we see a negative impact for boys at endline. This result is particularly surprising given the positive (insignificant) point estimate for primary completion, suggesting that boys are no less likely to complete primary school as a result of the program, but they do appear to be less likely to complete the first year of secondary. The point estimate is smaller but also negative for girls, and the difference is not statistically significant.

5.3.2 Household poverty

When we test for heterogeneous treatment effects by household poverty level (captured by an indicator for having below-median household consumption) in Table 6, we identify the largest gains in having ever attended school for the less-poor children. For the poorest children, point estimates are nearly zero at both midline and endline and statistically insignificant, while for the less-poor children, we identify a larger (compared to the overall effect of 4 percentage points found in Table 4), 5 percentage point increase in the likelihood of having ever attended school at midline that grows to a 6 percentage point increase by endline (significant at the 0.01 level whether using conventional or Romano-Wolf adjusted p-values). The results for less-poor children are statistically significantly different from those for poor children when using conventional p-values, though are short of conventional significance levels when using Romano-Wolf adjusted p-values.

Mirroring greater benefits for children in less-poor households, we find that at midline, the CCT program lowered the likelihood of having missed school in the last week by 6 percentage points for less-poor households (a result significant at the 0.10 level using the conventional p-value, but not significant at conventional levels when using the Romano-Wolf adjusted p-value); this stands in contrast to an opposite-signed (5 percentage point *increase* in the likelihood of missing school) that is statistically insignificant for the poorest households. At endline, we see a similar pattern of oppositely-signed point estimates for the less-poor vs. poorest, but the difference is statistically insignificant using Romano-Wolf adjusted p-values. We do not find significant differences in treatment effects for the poorest compared to the less-poor when using our administrative measure of attendance.

Overall, these suggestive results are consistent with CCT program transfers and other household resources being complements boosting attendance. The poorest households may lack resources that allow them to enroll and/or boost attendance (e.g., transfers may be diverted to food and not spent on items that promote attendance like shoes and payment of school fees).

[20] All four coefficient estimates and standard errors in Table 5, column 1 are identical when rounded to two decimal points, but differences in statistical significance across estimates are reported accurately and are due to rounding.

When our outcomes are completing primary school and completing at least one grade of secondary school—for which the CCT had statistically insignificant impacts for the full sample of children aged 14–21—we again do not identify any statistically significant impacts for either the poorest or less-poor half of households, nor do we identify statistically significant differences in the effects of transfers across the two groups.

Finally, we plot data on the treatment effect size by poverty quintile (Figure A5). These figures generally support the results from Table 6; that is, any increases in having ever attended school and declines in absenteeism are concentrated among less-poor children, while neither demographic benefits in terms of increased completion of the first grade of secondary school. One interesting finding in Figure A5, however, is that households in the bottom quintile of household consumption experience statistically significant improvements in primary completion at endline. This offers limited evidence that at least the CCT did not disadvantage the most vulnerable (in terms of poverty) groups on the outcome of primary completion, and possibly advantaged them.

5.3.3 Baseline exam performance

Our third vulnerability indicator is whether or not the child sat for and passed a Standard IV (i.e., 4th grade) exam at baseline. For those who did not sit for such an exam, this indicator takes a value of 0. In this analysis (Table 7), we find a mixed pattern regarding whether more versus less vulnerable children (at baseline) benefit more from the CCT program.

First considering our outcome related to whether children ever attended school, we find that those who had not sat for and passed a Standard IV exam at baseline benefited most. This is unsurprising, since those who had sat for and passed such an exam at baseline would all have "ever attended" as well, and so would have no room for growth on that variable. Beyond that finding, we observe at endline that the program increased the likelihood that parents reported a student was absent the previous week for those with stronger initial academic performance, although those results are not significant according to Romano-Wolf adjusted p-values.

However, we do find that it is those with stronger initial educational performance that were most likely to benefit from increased likelihood of having completed primary school. Specifically, the CCT program caused children who had sat and passed this exam at baseline to be 12 percentage points more likely to complete primary school at midline and 16 percentage points more likely to complete primary school at endline. Both of these results are statistically significant using either conventional of Romano-Wolf p-values. The CCT did not have statistically significant impacts on primary completion for those who had not passed this exam at baseline, and the point estimates are furthermore opposite-signed and smaller in magnitude. The CCT did not have a statistically significant effect on completion of the first grade of secondary school for either group (less- or more-vulnerable on this indicator). Having passed the Standard IV exam is a measure of either student academic performance, parent support and resources, or a combination of the two—which may be critical complements to transfers for outcomes like primary completion.

5.3.4 Heterogeneous effects using machine learning

We use a causal forest algorithm to complement our more traditional analysis of heterogeneous treatment effects (Wager and Athey, 2018). Researchers have increasingly turned to this and similar methods to analyze heterogeneity for randomized experiments due to their ability to sift through large amounts of information in a principled way. In short, a causal forest algorithm takes a dataset, including treatment assignment and covariates as inputs, and partitions the data along the covariates to maximize variance in treatment effects subject to a chosen parameter that limits the minimum cell size in which the data can be partitioned. Repeating this process thousands of times allows the algorithm to predict a conditional average treatment effect (CATE) for all individuals in the dataset. We use the "grf" package in R to implement the causal forest algorithm (Tibshirani et al., 2020). Appendix A.2 describes the causal forest procedure in greater detail. We consider a wide range of covariates including the three theoretically motivated covariates discussed in previous sections and indicators capturing the range of heterogeneous treatment effects identified by two reviews—Baird et al. (2014) and Bastagli et al. (2016) (see Table 8).

One approach to evaluating which dimensions of heterogeneity are most important is to examine which characteristics the algorithm finds are most important in terms of determining splits. The grf package provides a calculated value for each covariate and outcome which captures the weighted sum of the number of times on which that covariate was split at each depth of the forest. Figure 1 presents these results, displaying the five most important covariates for each of the five key outcomes we presented in Table 4. For our outcome indicating having ever attended school—arguably our most important outcome, as it is the one most robustly impacted by treatment—the most important covariate was a dummy for data on the first grade of secondary being completed taking on a value of missing, followed by age, household head age, and total household consumption. Household poverty was also identified as an important predictor of treatment effects on attendance in the traditional, post-hoc analysis described in section 5.3.2. However, using the machine learning algorithm identifies a number of other important predictors with less theoretical motivation in the literature. Looking across the four other outcomes considered in Table 4 (appearing in panels B through E of Figure 1), it is clear that other continuous measures—such as village population, distance to school, and total number of household members—are important for determining splits in the causal forest. Among these, it is interesting to note that distance to school appears important for determining heterogeneity in treatment effects for school progression and completion outcomes (measured for older children only), but less so for attendance outcomes (measured for all children aged 6–21).

Across the five outcomes considered in Figure 1, we find five covariates that are important for explaining variation in treatment effect sizes for four or more outcomes: total household consumption, household head age, child age, village population, and distance to school. Taking these five covariates, there are 10 combinations of two covariates. Figure 2 then considers each of these 10 combinations, considering how CATEs for our ever attended school outcome vary across quintiles of each covariate. This produces a 25-cell heat map for each covariate combination. Yellow indicates larger CATEs while purple indicates the smallest CATEs. Examining this heat map, we observe which combinations of covariates matter—and get insights into the direction of effects.

Overall, we identify fairly little heterogeneity in CATEs for our ever attended outcome; they range from 0.008 to 0.012. There are larger treatment effects in poorer households with older household heads. Additionally, we identify larger treatment effects for older children who live closer to schools, as well as for children in households with relatively old heads who live close to schools. There are few clear patterns when it comes to village population, however. Overall, the machine learning results suggest that factors less motivated by theory may be important for predicting treatment effects of cash transfers on schooling outcomes.

6 Discussion

Our results contribute to a large literature on cash transfers and education. Two recent reviews – Baird et al. (2014) and Bastagli et al. (2016) – cite more than one hundred separate studies between them of the impact of cash transfers on education. Our average impacts are broadly consistent with the literature, which shows positive impacts on school participation. We add to that our null results on primary completion rates, of which neither earlier review makes explicit mention.

In terms of heterogeneous treatment effects, Baird et al. (2014) report comparable but slightly larger effects on attendance for girls, although the differences are unlikely to be statistically significant. These are broadly consistent with our results of similar point estimates for girls and boys, but with statistical significance concentrated among girls—suggesting greater precision for the latter estimates (Table 5).

Only 29 percent of the 101 earlier studies separate impacts for the poorest and the less-poor (Table 8). Cash transfers are almost always targeted to households that are poor in general, so these distinctions are within that context. We find results concentrated among the less-poor of our sample. Barrera-Osorio et al. (2011) find stronger enrollment effects for the poorest students in Colombia. In Brazil, Cardoso and Souza (2009) report the strongest impacts on school enrollment for children with uneducated parents. The same is true for attendance and enrollment among secondary school girls in Cambodia (Filmer and Schady, 2008). A study in Kenya finds positive impacts on school completion, concentrated among the poorest (Merttens et al., 2013). In China, alternatively, the impact of a cash transfer program on dropout rates is smaller among the poorest (Mo et al., 2013). Thus, our results and others suggest that the relationship between cash transfers, poverty, and schooling outcomes may not be straightforward or consistent across contexts.

Even fewer previous studies examine the impacts of cash transfers for children with differing initial school access (20 percent) or performance (7 percent). In Burkina Faso, conditional cash transfers had stronger impacts for lower performing students than did unconditional cash transfers (Akresh et al., 2013). An evaluation in Ghana found strong impacts for boys with lower measured initial ability (de Groot et al., 2015). Our results are consistent with these findings.

While one mechanism through which cash transfers can affect educational outcomes is that households actually use the cash to pay for educational expenses, this does not seem to be the case here. Treatment and comparison households experienced similar increases in educational expenditures between baseline and endline, as outlined in Section 2.4. But despite being conditioned on the fulfillment of certain activities, the actual spending of the money is unconditional, so either households may have used the money for non-educational goods that still facilitate schooling (e.g., shoes or healthcare) or they may have simply increased efforts and attention towards schooling in order to ensure receipt of the transfers, overcoming a non-monetary constraint.

7 Conclusion

In this paper, we identify the impact of a community-based conditional cash transfer program – targeted to poor children in Tanzania – on those children's educational engagement. We find positive average effects on children having ever attended school, likely driven by higher rates of age-appropriate enrollment for school-age children. We go on to show that some other indicators of vulnerability – like gender – do not show differential impacts, whereas others – like household wealth and initial educational outcomes – do show differential impacts.

We observe children for 2.75 years after transfers, which is longer than the vast majority of impact evaluations of education interventions. McEwan (2015) shows that most education evaluations examine impacts after one year or less. However, the ultimate objective is not only to keep children in school in the short run, but to help them achieve more total schooling and better post-school life outcomes. Longer-run data on the impact of cash transfers on children with a variety of indicators of vulnerability would be valuable, particularly with sample sizes providing sufficient statistical power to examine these heterogeneities. Furthermore, this work does not evaluate the impact of cash transfers for non-beneficiaries within the same village, and existing work suggests some evidence both for positive economic impacts (Egger et al., 2019; Handa et al., 2019) and adverse psychological impacts (Haushofer et al., 2019). Future work can further document spillovers to a variety of vulnerable groups within villages. Additionally, while our pattern of relative significance remains after we adjust for multiple hypothesis testing, some statistical significance disappears, potentially due to limited samples of different heterogeneous groups.

Ultimately, cash transfers are not inherently designed to address all vulnerabilities, which means that either they may need to be adapted to benefit children with vulnerabilities beyond poverty or complemented with other policy programs to ensure that all children have access to investments in quality human capital.

References

Acerenza, S. and N. Gandelman (2019). Household education spending in Latin America and the Caribbean: Evidence from income and expenditure surveys. *Education Finance and Policy 14*(1), 61–87.

Akresh, R., D. De Walque, and H. Kazianga (2013). Cash transfers and child schooling: Evidence from a randomized evaluation of the role of conditionality. *World Bank Policy Research Working Paper* (6340).

Alderman, H. and E. M. King (1998). Gender differences in parental investment in education. *Structural Change and Economic Dynamics 9*(4), 453–468.

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

Araujo, M. C., M. Bosch, and N. Schady (2019). Can cash transfers help households escape an intergenerational poverty trap? In C. B. Barrett, J.-P. Chavas, and M. R. Carter (Eds.), *The Economics of Poverty Traps*. Chicago, IL: University of Chicago Press.

Aslam, M. and G. G. Kingdon (2008). Gender and household education expenditure in Pakistan. *Applied Economics 40*(20), 2573–2591.

Athey, S. and G. Imbens (2016). Recursive partitioning for heterogeneous causal effects. *Proceedings of the National Academy of Sciences 113*(9).

Baird, S., F. H. Ferreira, B. Özler, and M. Woolcock (2014). Conditional, unconditional and everything in between: A systematic review of the effects of cash transfer programmes on schooling outcomes. *Journal of Development Effectiveness 6*(1), 1–43.

Baird, S., C. McIntosh, and B. Özler (2013, October). The regressive demands of demand-driven development. *Journal of Public Economics 106*, 27–41.

Banerjee, A., R. Banerji, J. Berry, E. Duflo, H. Kannan, S. Mukerji, M. Shotland, and M. Walton (2017). From proof of concept to scalable policies: Challenges and solutions, with an application. *Journal of Economic Perspectives 31*(4), 73–102.

Barrera-Osorio, F., M. Bertrand, L. Linden, and F. Perez-Calle (2011). Improving the design of conditional transfer programs: Evidence from a randomized education experiment in Colombia. *American Economic Journal: Applied Economics 3*, 167195.

Barrera-Osorio, F., L. L. Linden, and J. E. Saavedra (2019). Medium- and long-term educational consequences of alternative conditional cash transfer designs: Experimental evidence from Colombia. *American Economic Journal: Applied Economics 11*(3), 54–91.

Bastagli, F., J. Hagen-Zanker, L. Harman, V. Barca, G. Sturge, T. Schmidt, and L. Pellerano

(2016). Cash transfers: What does the evidence say? A rigorous review of programme impact and of the role of design and implementation features.

Battaile, W. G. (2020). What does Tanzania's move to lower-middle income status mean? *Africa Can End Poverty Blog*.

Becker, G. S. and N. Tomes (1976). Child endowments and the quantity and quality of children. *Journal of political Economy 84*(4, Part 2), S143–S162.

Belghith, y. N. B. H., W. Karamba, E. Talbert, and P. de Boisseson (2019). Tanzania: Mainland poverty assessment.

Buhl-Wiggers, J., S. Jones, and S. Thornton (2021). Boys lagging behind: Unpacking gender differences in academic achievement across East Africa. *International Journal of Educational Development 83*.

Cardoso, E. and A. P. F. de Souza (2009). The impact of cash transfers on child labor and school enrollment in Brazil. In P. Orazem, G. Sedlacek, and Z. Tzannatos (Eds.), *Child Labor and Education in Latin America*. New York: Palgrave Macmillan.

Cho, H., R. C. Ryberg, K. Hwang, L. D. Pearce, and B. J. Iritani (2017). A school support intervention and educational outcomes among orphaned adolescents: Results of a cluster randomized controlled trial in Kenya. *Prevention Science 18*, 943954.

Clarke, D., J. P. Romano, and M. Wolf (2020). The romano–wolf multiple-hypothesis correction in stata. *The Stata Journal 20*(4), 812–843.

de Groot, R., S. Handa, M. Park, R. O. Darko, I. Osei-Akoto, G. Bhalla, and L. P. Ragno (2015). Heterogeneous impacts of an unconditional cashtransfer programme on schooling: Evidence from the Ghana LEAP Programme. *Office of Research Innocenti Working Paper 2015-10*.

Dizon-Ross, R. (2019). Parents' beliefs about their children's academic ability: Implications for educational investments. *American Economic Review 109*(8), 2728–65.

Dizon-Ross, R. and S. Jayachandran. Dads and daughters: Disentangling altruism and investment motives for spending on children. Technical report.

Duflo, A., J. Kiessel, and A. Lucas (2020). External validity: Four models of improving student achievement. *National Bureau of Economic Research (NBER)*.

Egger, D., J. Haushofer, E. Miguel, P. Niehaus, and M. W. Walker (2019). General equilibrium effects of cash transfers: Experimental evidence from Kenya. *National Bureau of Economic Research (NBER)*.

Evans, D. K., M. Akmal, and P. Jakiela (2021). Gender gaps in education: The long view. *IZA Journal of Development and Migration*.

Evans, D. K., B. Holtemeyer, and K. Kosec (2017). Cash Transfers and Health: Evidence from Tanzania. *The World Bank Economic Review*.

Evans, D. K., B. Holtemeyer, and K. Kosec (2019a). Cash transfers and health: Evidence from Tanzania. *World Bank Economic Review 33*(2), 394–412.

Evans, D. K., B. Holtemeyer, and K. Kosec (2019b, February). Cash transfers increase trust in local government. *World Development 114*, 138–155.

Evans, D. K. and K. Kosec (2020). Do cash transfers reduce trust and informal transfers within communities? *IFPRI Discussion Paper 01994*.

Evans, D. K. and A. Mendez Acosta (2021). How to measure student absenteeism in low- and middle-income countries. *Center for Global Development Working Paper 600.*

Fazzio, I., A. Eble, R. L. Lumsdaine, P. Boone, B. Bouy, P.-T. J. Hsieh, C. Jayanty, S. Johnson, and A. F. Silva (2020). Large learning gains in pockets of extreme poverty: Experimental evidence from Guinea Bissau. *NBER Working Paper 27799.*

Filmer, D. and N. Schady (2008). Getting girls into school: Evidence from a scholarship program in Cambodia. *Economic Development and Cultural Change 56*(3), 581–617.

Hagen-Zanker, J., F. Bastagli, L. Harman, V. Barca, G. Sturge, and T. Schmidt (2016). Understanding the impact of cash transfers: The evidence. Technical report, Overseas Development Institute.

Hallfors, D., H. Cho, S. Rusakaniko, B. Iritani, J. Mapfumo, and C. Halpern (2011). Supporting adolescent orphan girls to stay in school as HIV risk prevention: Evidence from a randomized controlled trial in Zimbabwe. *American Journal of Public Health 101*, 1082–1088.

Handa, S., L. Natali, D. Seidenfeld, G. Tembo, and B. Davis (2019). Can unconditional cash transfers raise long-term living standards? Evidence from Zambia. *Journal of Development Economics 133*, 42–65.

Haushofer, J., J. Reisinger, and J. Shapiro (2019). Is your gain my pain? Effects of relative income and inequality on psychological well-being. *Working Paper*.

Hickey, S., T. Lavers, M. Nio-Zaraza, and J. Seekings (Eds.) (2019, November). *The Politics of Social Protection in Eastern and Southern Africa.* WIDER Studies in Development Economics. Oxford, New York: Oxford University Press.

Masterson, T. (2012). An empirical analysis of gender bias in education spending in Paraguay. *World Development 40*(3), 583–593.

McEwan, P. (2015). Improving learning in primary schools of developing countries: A meta-analysis of randomized experiments. *Review of Educational Research 85.*

Merttens, F., A. Hurrell, M. Marzi, R. Attah, M. Farhat, A. Kardan, and I. MacAuslan (2013).

Kenya hunger safety net programme monitoring and evaluation component impact evaluation final report: 2009 to 2012.

Mo, D., L. Zhang, H. Yi, R. Luo, S. Rozelle, and C. Brinton (2013). School dropouts and conditional cash transfers: Evidence from a randomised controlled trial in rural China's junior high schools. *Journal of Development Studies 49*(2), 190–207.

Molina Millán, T., T. Barham, K. Macours, and J. A. Maluccio (2019). Long-term impacts of conditional cash transfers: Review of the evidence. *World Bank Research Observer 34*(1), 119–159.

Molina Millán, T., K. Macours, J. A. Maluccio, and L. Tejerina (2020). Experimental long-term effects of early-childhood and school-age exposure to a conditional cash transfer program. *Journal of Development Economics 143*.

Moodley, J. and L. Graham (2015). The importance of intersectionality in disability and gender studies. *Agenda: Empowering Women for Gender Equity*.

Ogundari, K. and A. Abdulai (2014). Determinants of household's education and healthcare spending in Nigeria: Evidence from survey data. *African Development Review 26*(1), 1–14.

Pasqua, S. (2005). Gender bias in parental investments in childrens education: A theoretical analysis. *Review of Economics of the Household 3*(3), 291–314.

Psacharopoulos, G. and H. A. Patrinos (2018). Returns to investment in education: A decennial review of the global literature. *Education Economics 26*(5), 445–458.

Psaki, S., N. Haberland, B. Mensch, L. Woyczynski, and E. Chuang (2022). Policies and interventions to remove gender-related barriers to girls' school participation and learning in low- and middle-income countries: A systematic review of the evidence. *Campbell Systematic Reviews 18*(1), e1207.

Sabates, R., P. Rose, B. Alcott, and M. Delprato (2020). Assessing cost-effectiveness with equity of a programme targeting marginalised girls in secondary schools in Tanzania. *Journal of Development Effectiveness*.

Shupavu, B. (2020). Tanzania's school system: An overview. *AfricaAid*.

Thomas, T., Y. Ahmed, M. Tan, and E. L. Grigorenko (2020). Cognitive and educational interventions for orphans and vulnerable children affected by HIV/AIDS: A metaanalysis. *Child Development*.

Tibshirani, J., S. Athey, D. Friedberg, V. Hadad, D. Hirshberg, L. Miner, E. Sverdrup, S. Wager, and M. Wright (2020). grf package.

UNESCO (2022). *Leave no child behind: global report on boys disengagement from education.* UNESCO.

Uwezo (2019). Are our children learning? uwezo uganda eighth learning assessment report.

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

Welmond, M. J. and L. Gregory (2022). *Educational Underachievement Among Boys and Men*. World Bank.

Wongmonta, S. and P. Glewwe (2017). An analysis of gender differences in household education expenditure: The case of Thailand. *Education Economics 25*(2), 183–204.

World Bank (2018). *World Development Report 2018: Learning to Realize Educations Promise*. World Bank.

World Bank (2019a). *Ending Learning Poverty: What Will It Take?* World Bank.

World Bank (2019b). World development indicators.

Table 1: Timeline for Implementation of CCT and Accompanying Impact Evaluation

Timing	Activity
November 2007 - September 2008	Program Design (completion of Operational Manual, set up of MIS, preparation of guidelines, forms, and materials for training activities)
September - November 2008	Sensitization at regional, district, ward, and community levels
October - November 2008	Targeting activities (field data collection, data entry, and community validation of beneficiaries)
October - November 2008	Training of district officers and community management committees on the targeting process
January - May 2009	Baseline survey
September - October 2009	Enrollment of beneficiaries
January 2010	First payments made to beneficiary households
November 2010 - February 2011	Community Scorecard Exercise
July - September 2011	Midline survey & first round of focus group interviews
August - October 2012	Endline survey
July - August 2013	Second round of qualitative data collection, including in-depth and focus group interviews

Table 2: Balance Across Groups at Baseline

	Treatment (T)		Control (C)		Difference (T-C)	
	$\beta_0+\beta_1$	N	β_0	N	β_1	S.E.
	(1)	(2)	(3)	(4)	(5)	(6)
Individuals aged 6-21						
Age (years)	12.41	1243	12.55	1181	-0.15	0.18
Female	0.47	1243	0.47	1181	0.00	0.02
Either parent is deceased	0.28	1108	0.31	1041	-0.03	0.03
Neither parent in HH	0.46	1036	0.49	954	-0.03	0.04
Passed primary exam	0.29	1243	0.31	1181	-0.03	0.02
Ever attended	0.81	1243	0.83	1181	-0.02	0.02
Missed school last week	0.45	1243	0.42	1181	0.03	0.03
F-test joint sig.	0.45					
Individuals aged 14-21						
Completed primary	0.48	445	0.52	441	-0.04	0.04
Completed first grade of secondary	0.07	451	0.15	441	-0.07**	0.03
F-test joint sig.	3.19**					
Households						
Household head has no education	0.55	547	0.60	516	-0.05	0.04
Less than median HH consumption	0.33	547	0.32	517	0.01	0.03
Head says school quality good or excellent	0.82	546	0.85	517	-0.03	0.03
Total HH members (#)	5.21	547	5.31	517	-0.10	0.16
Total HH members 0-5 (#)	0.66	547	0.66	517	0.00	0.06
Total HH members 7-15 (#)	1.54	547	1.52	517	0.02	0.08
Total HH members 60+ (#)	1.14	547	1.30	517	-0.16**	0.07
F-test joint sig.	1.43					
Villages						
Number of HH 2009	868.02	39	1091.42	39	-223.40	344.08
Village houses ward	0.27	40	0.43	40	-0.16	0.11
VEO lives in village	0.83	40	0.70	40	0.13	0.09
Num. general village meetings held	3.15	40	3.35	40	-0.21	0.24
Poverty rank (TASAF)	1.19	40	1.13	40	0.06	0.10
PMT Poverty score (TASAF)	7575.99	40	7398.65	40	177.34	236.07
F-test joint sig.	1.21					

Source: Authors' calculations based on 2009 household survey data.

Note: Table presents results from an ordinary least squares regression of each covariate separately on treatment, including district dummies and standard errors clustered at the village. Whether child missed school in last schooling week includes those who missed for a reason other than public holiday, school closure or teacher absence. Consumption measured as total annualized food consumption value plus non-food expenditure. Variables labeled TASAF are taken from proxy means test conducted prior to randomization; the PMT (proxy means test) poverty score is estimated from household consumption and assets including roofing, source of power and the presence of appliances, and the poverty rank is a score (0 to 5) assigned to households calculated based on the number of would-be benficiaries, status of the household head and household amenities.

*** significant at the 1% level

** significant at the 5% level

* significant at the 10% level.

Table 3: Attrition

	Household is in ... survey				Individual 6-21 is in ... survey			
	Midline		Endline		Midline		Endline	
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Treatment village	0.002	0.088	-0.004	0.107	0.001	0.186	-0.002	0.231
	(0.003)	(0.077)	(0.004)	(0.081)	(0.010)	(0.162)	(0.012)	(0.188)
Head no ed.		0.001		0.011*		0.024*		0.029*
		(0.001)		(0.006)		(0.014)		(0.017)
Treat * Head no ed.		0.004		-0.006		-0.000		0.007
		(0.005)		(0.008)		(0.017)		(0.021)
Low wealth		0.005		0.020*		0.001		-0.003
		(0.007)		(0.011)		(0.022)		(0.029)
Treat * Low wealth		0.003		-0.015		0.002		0.028
		(0.008)		(0.012)		(0.026)		(0.034)
Head is male		-0.003		0.005		0.007		0.004
		(0.003)		(0.008)		(0.015)		(0.015)
Treat * Head is male		-0.004		-0.012		-0.005		0.011
		(0.006)		(0.010)		(0.020)		(0.026)
Head age		0.001		0.001		0.003		0.004
		(0.001)		(0.001)		(0.002)		(0.003)
Treat * Head age		-0.003		-0.003		-0.008*		-0.009*
		(0.002)		(0.003)		(0.004)		(0.005)
Head age sq.		-0.000		-0.000		-0.000		-0.000
		(0.000)		(0.000)		(0.000)		(0.000)
Treat * Head age sq.		0.000		0.000		0.000*		0.000*
		(0.000)		(0.000)		(0.000)		(0.000)
Female						0.006		0.009
						(0.011)		(0.012)
Treat * Female						-0.001		0.012
						(0.018)		(0.021)
Ind. age						-0.039***		-0.040***
						(0.011)		(0.012)
Treat * Ind. age						0.005		0.001
						(0.015)		(0.016)
Ind. age sq.						0.002***		0.002***
						(0.000)		(0.000)
Treat * Ind. age sq.						-0.000		-0.000
						(0.001)		(0.001)
Constant	0.002	-0.014	0.008**	-0.033	0.040***	0.162**	0.057***	0.163*
	(0.002)	(0.014)	(0.003)	(0.031)	(0.007)	(0.079)	(0.009)	(0.091)
Observations	1064	1063	1064	1063	2424	2422	2424	2422
R-squared	0.003	0.019	0.011	0.025	0.008	0.035	0.016	0.042
F-stat. joint signif. of interactions		0.475		0.788		0.546		0.704

Source: Authors' calculations based on 2009, 2011 and 2012 household survey data.
Note: Table presents results from an OLS regression of a dummy variable to indicate attrition in the corresponding survey round on covariates listed in table, with each column showing results for a separate regression. An individual is considered to have attrit if their unique ID in the dataset does not appear in the corresponding survey round. Standard errors clustered by village. Household is low wealth if it had less than median household consumption at baseline.

*** significant at the 1% level
** significant at the 5% level
* significant at the 10% level.

Table 4: Conditional Cash Transfers Impacts on Schooling Outcomes

	Ever Attended (age 6-21) (1)	Missed Last Week (age 6-21) (2)	School Attendance (age 6-21) (3)	Primary Completed (age 14-21) (4)	First Grade Sec. Completed (age 14-21) (5)
Treatment X Midline	0.04**	-0.03		0.01	-0.05
	(0.02)	(0.03)		(0.04)	(0.05)
	[0.01]	[0.54]		[0.73]	[0.50]
Treatment X Endline	0.04**	-0.02	0.03	0.03	-0.09
	(0.02)	(0.04)	(0.04)	(0.04)	(0.06)
	[0.01]	[0.67]	[0.54]	[0.67]	[0.15]
Midline	0.08***	0.02		0.36***	0.23***
	(0.01)	(0.03)		(0.05)	(0.04)
Endline	0.11***	0.04		0.54***	0.33***
	(0.02)	(0.04)		(0.05)	(0.06)
Observations	8963	7670	1943	3284	2992
Baseline mean	0.82	0.44		0.50	0.11

Source: Authors' calculations based on 2009, 2011, and 2012 household survey data.

Note: Table presents estimates of effect of CCT at midline and endline. Each column is a separate regression. The analytic sample is children who were 6 to 21 or 14 to 21 at the time of survey. Dependent variables are a dummy to indicate whether child has ever attended school (self-reported), whether child missed school in last schooling week for a reason other than public holiday, school closure or teacher absence (self-reported), whether child was recorded as attending more than 80% of open school days in May 2012 school register (endline for primary schools only), whether child completed primary (self-reported), and whether child completed first grade of secondary (self-reported). Romano-Wolf adjusted p-values are shown in brackets and standard errors, clustered by village, are shown in parentheses. All regressions control for the baseline value of each imbalanced variable in Table 2, a dummy to indicate this variable was missing and interactions of each with period. Note that asterisks refer to standard (non-adjusted) p-values.

*** significant at the 1% level

** significant at the 5% level

* significant at the 10% level.

Table 5: Heterogeneous Effects of CCT by Gender

	Ever Attended (age 6-21) (1)	Missed Last Week (age 6-21) (2)	School Attendance (age 6-21) (3)	Primary Completed (age 14-21) (4)	First Grade Sec. Completed (age 14-21) (5)
Treatment X Midline X Female	0.04**	-0.03		-0.03	0.01
	(0.02)	(0.05)		(0.06)	(0.06)
	[0.04]	[0.98]		[0.99]	[1.00]
Treatment X Midline X Male	0.04	-0.02		0.04	-0.09
	(0.02)	(0.04)		(0.06)	(0.06)
	[0.43]	[0.99]		[0.98]	[0.48]
Treatment X Endline X Female	0.04*	-0.00	0.04	0.00	-0.02
	(0.02)	(0.05)	(0.05)	(0.06)	(0.08)
	[0.13]	[1.00]	[0.94]	[1.00]	[1.00]
Treatment X Endline X Male	0.04	-0.04	0.03	0.04	-0.13**
	(0.02)	(0.05)	(0.04)	(0.06)	(0.06)
	[0.36]	[0.94]	[0.98]	[0.98]	[0.10]
Observations	8963	7670	1943	3284	2992
Baseline mean female	0.81	0.44		0.54	0.10
Baseline mean male	0.83	0.44		0.47	0.12
Midline p-value of difference	0.97	0.87		0.39	0.20
Midline RW-adjusted p-value	1.00	1.00		0.95	0.69
Endline p-value of difference	0.92	0.49	0.79	0.62	0.16
Endline RW-adjusted p-value	1.00	0.98	1.00	0.99	0.59

Source: Authors' calculations based on 2009, 2011, and 2012 household survey data.
Note: Table presents estimates of effects of CCT at midline and endline. Estimates for midline and endline coefficients are not shown. The analytic sample is children who were 6 to 21 or 14 to 21 at the time of survey. Dependent variables are a dummy to indicate whether child has ever attended school (self-reported), whether child missed school in last schooling week for a reason other than public holiday, school closure or teacher absence (self-reported), whether child was recorded as attending more than 80% of open school days in May 2012 school register (endline for primary schools only), whether child completed primary (self-reported), and whether child completed first grade of secondary (self-reported). All regressions control for the baseline value of each imbalanced variable in Table 2, a dummy to indicate this variable was missing and interactions of each with period. Romano-Wolf adjusted p-values are shown in brackets and standard errors, clustered by village, are shown in parentheses. Note that asterisks refer to standard (non-adjusted) p-values. Although all four coefficient estimates and standard errors Column 1 are identical when rounded to two decimal points, differences in statistical significance across estimates are accurately reported and are due to rounding.

*** significant at the 1% level
** significant at the 5% level
* significant at the 10% level.

Table 6: Heterogeneous Effects of CCT by Wealth

	Ever Attended (age 6-21) (1)	Missed Last Week (age 6-21) (2)	School Attendance (age 6-21) (3)	Primary Completed (age 14-21) (4)	First Grade Sec. Completed (age 14-21) (5)
Treatment X Midline X Poorest	0.00	0.05		-0.02	-0.04
	(0.03)	(0.06)		(0.10)	(0.07)
	[1.00]	[0.91]		[1.00]	[0.99]
Treatment X Midline X Less Poor	0.05***	-0.06*		0.01	-0.05
	(0.02)	(0.03)		(0.04)	(0.05)
	[0.00]	[0.30]		[1.00]	[0.86]
Treatment X Endline X Poorest	-0.006	0.07	0.01	-0.00	-0.14
	(0.033)	(0.07)	(0.06)	(0.11)	(0.09)
	[1.00]	[0.90]	[1.00]	[1.00]	[0.47]
Treatment X Endline X Less Poor	0.06***	-0.06	0.04	0.03	-0.07
	(0.02)	(0.04)	(0.04)	(0.05)	(0.06)
	[0.00]	[0.53]	[0.78]	[0.98]	[0.78]
Observations	8963	7670	1943	3284	2992
Baseline mean poorest	0.75	0.47		0.35	0.07
Baseline mean less poor	0.84	0.43		0.53	0.12
Midline p-value of difference	0.10	0.08		0.78	0.89
Midline RW-adjusted p-value	0.34	0.29		1.00	1.00
Endline p-value of difference	0.06	0.09	0.66	0.80	0.46
Endline RW-adjusted p-value	0.20	0.30	1.00	1.00	0.96

Source: Authors' calculations based on 2009, 2011, and 2012 household survey data.

Note: Table presents estimates of effects of CCT at midline and endline. Poorest households are those in which total annual consumption was less than or equal to the median at baseline, while less poor households are those in which total annual consumption was greater than the median. Estimation for midline and endline coefficients are not shown. The analytic sample is children who were 6 to 21 (or 14 to 21, as indicated) at the time of survey. Dependent variables are a dummy to indicate whether child has ever attended school (self-reported), whether child missed school in last schooling week for a reason other than public holiday, school closure or teacher absence (self-reported), whether child was recorded as attending more than 80% of open school days in May 2012 school register (endline for primary schools only), whether child completed primary (self-reported), and whether child attended first grade of secondary (self-reported). All regressions control for the baseline value of each imbalanced variable in Table 2, a dummy to indicate this variable was missing and interactions of each with period. Romano-Wolf adjusted p-values are shown in brackets and standard errors, clustered by village, are shown in parentheses. Note that asterisks refer to standard (non-adjusted) p-values.

*** significant at the 1% level

** significant at the 5% level

* significant at the 10% level.

Table 7: Heterogeneous Effects of CCT by Exam Performance

	Ever Attended (age 6-21) (1)	Missed Last Week (age 6-21) (2)	School Attendance (age 6-21) (3)	Primary Completed (age 14-21) (4)	First Grade Sec. Completed (age 14-21) (5)
Treatment X Midline X Not Passed	0.06**	-0.05		-0.06	-0.02
	(0.02)	(0.04)		(0.04)	(0.04)
	[0.01]	[0.35]		[0.35]	[0.96]
Treatment X Midline X Passed	-0.002	0.07		0.12*	-0.05
	(0.001)	(0.07)		(0.06)	(0.06)
	[0.32]	[0.66]		[0.08]	[0.80]
Treatment X Endline X Not Passed	0.06**	-0.05	0.02	-0.06	-0.06
	(0.03)	(0.04)	(0.04)	(0.05)	(0.05)
	[0.02]	[0.54]	[0.93]	[0.59]	[0.41]
Treatment X Endline X Passed	-0.002	0.13*	0.01	0.16***	-0.08
	(0.001)	(0.07)	(0.03)	(0.06)	(0.07)
	[0.35]	[0.18]	[0.96]	[0.00]	[0.58]
Observations	8963	7670	1943	3284	2992
Baseline mean not passed	0.74	0.52		0.66	0.08
Baseline mean passed	1.00	0.24		0.36	0.13
Midline p-value of difference	0.01	0.13		0.01	0.62
Midline RW-adjusted p-value	0.01	0.31		0.01	0.96
Endline p-value of difference	0.02	0.03	0.69	0.00	0.82
Endline RW-adjusted p-value	0.01	0.03	0.96	0.00	0.96

Source: Authors' calculations based on 2009, 2011, and 2012 household survey data.
Note: Table presents estimates of effects of CCT at midline and endline. Passed refers to individuals who passed a Primary (either std. IV or VII) exam at baseline, regardless of whether they sat the exam. Not passed similarly refers to individuals who had not passed a Primary exam at baseline. Estimation for midline and endline coefficients are not shown. The analytic sample is children who were 6 to 21 (or 14 to 21, as indicated) at the time of survey. Dependent variables are a dummy to indicate whether child has ever attended school (self-reported), whether child missed school in last schooling week for a reason other than public holiday, school closure or teacher absence (self-reported), whether child was recorded as attending more than 80% of open school days in May 2012 school register (endline for primary schools only), whether child completed primary (self-reported), and whether child completed first grade of secondary (self-reported). All regressions control for the baseline value of each imbalanced variable in Table 2, a dummy to indicate this variable was missing and interactions of each with period. Romano-Wolf adjusted p-values are shown in brackets and standard errors, clustered by village, are shown in parentheses. Note that asterisks refer to standard (non-adjusted) p-values.

*** significant at the 1% level

** significant at the 5% level

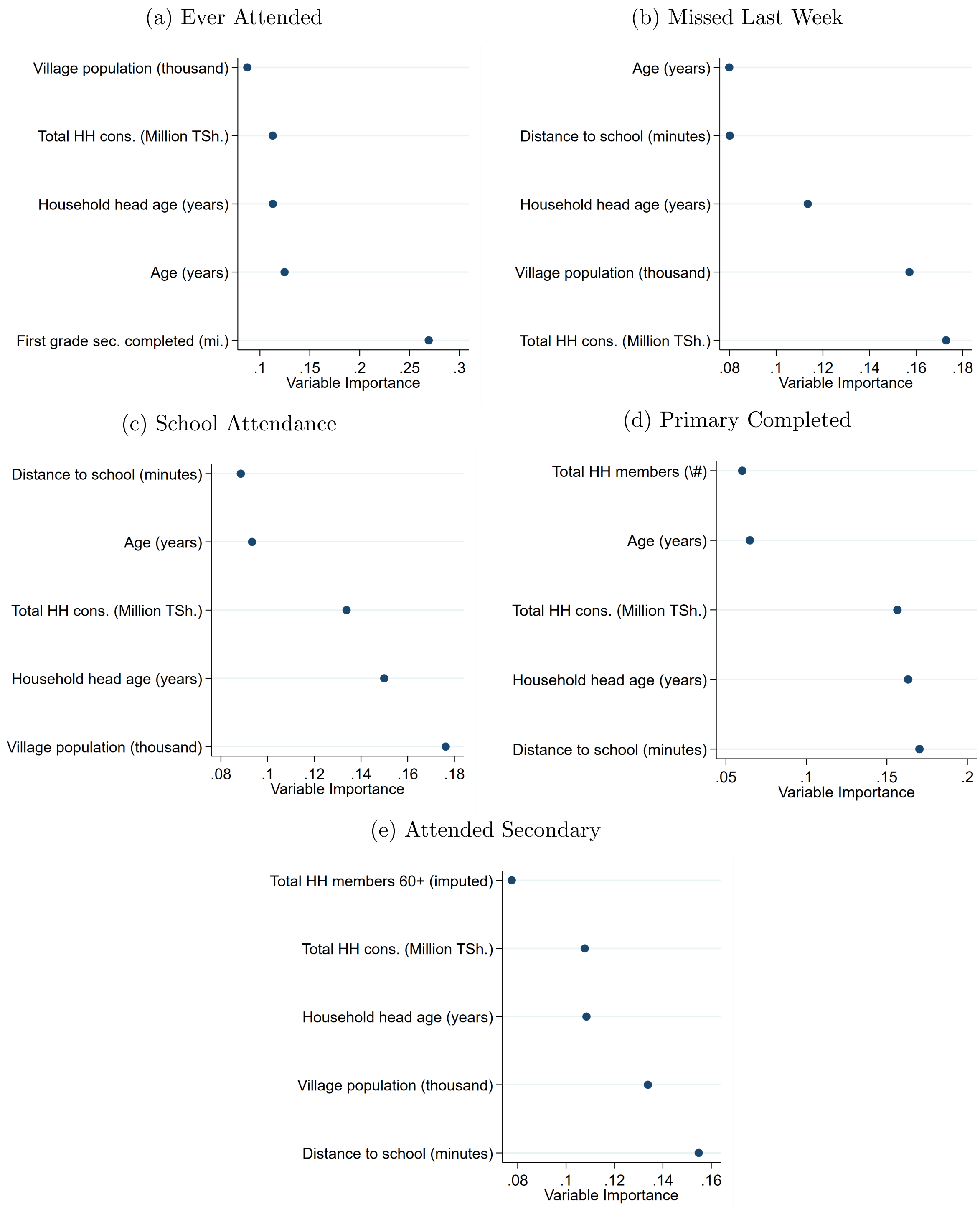
* significant at the 10% level.

Table 8: Aspects of Heterogeneity Analyzed in 101 Earlier Studies of Cash Transfers on Educational Outcomes

Proportion of studies that report results separately by...	
Gender	71%
Relative poverty	29%
Orphanhood	1%
Foster status	1%
Baseline school performance	7%
Baseline school access	20%
Child age	70%
Urban / rural	23%
Other vulnerabilities	28%

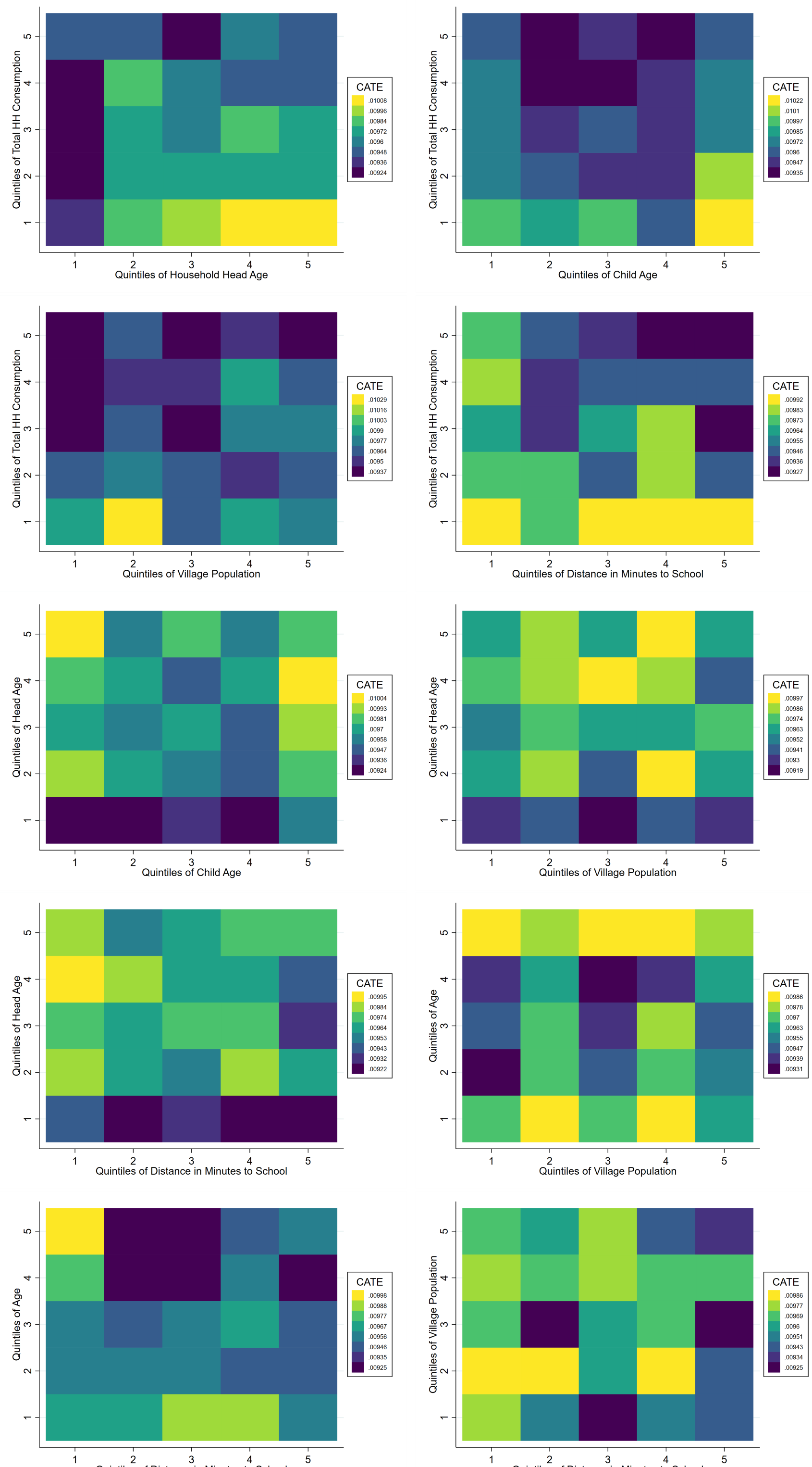
Note: The 101 studies include all studies in English referenced in Baird et al. (2014b) and Bastagli et al. (2016). (Three studies were omitted because they were in Spanish or Portuguese.) "Other vulnerabilities" include households with a single parent, households in indigenous areas, parent education or employment, race, etc. The full list of included and excluded studies, together with the analysis of heterogeneity, is available at https://www.cgdev.org/sites/default/files/2022-07/Evans-heterogeneity-cash-transfer-data-revised-format.xlsx.

Figure 1: Causal Forest Variable Importance



Note: Plots show variable importance, or the weighted sum of the number of times a variable was split on in using the causal forest algorithm, for each outcome.

Figure 2: Heat Maps of Conditional Average Treatment Effect (CATE) for Ever Attend



Note: Plots show heat maps of conditional average treatment effects (CATEs) on ever attended, by continuous variables used in machine learning algorithm.

A Appendix

Table A1: Replication of Main Effects with Alternative Age Group for Attendance Variables

	Ever Attended (age 7-15) (1)	Missed Last Week (age 7-15) (2)	School Attendance (age 7-15) (3)
Treatment X Midline	0.04*	-0.01	
	(0.02)	(0.03)	
	[0.06]	[0.95]	
Treatment X Endline	0.04	-0.01	0.02
	(0.02)	(0.04)	(0.04)
	[0.13]	[0.95]	[0.95]
Midline	0.00	-0.04	
	(0.00)	(0.04)	
Endline	0.13***	-0.05	
	(0.02)	(0.05)	
Observations	5597	5070	1488
Baseline mean	0.86	0.30	

Source: Authors' calculations based on 2009, 2011, and 2012 household survey data.

Note: Table presents estimates of effect of CCT at midline and endline. Analysis replicates Table 4 exactly for all outcomes except changing the age group presented. Each column is a separate regression. The analytic sample is children who were 7 to 15 at the time of survey. Dependent variables are a dummy to indicate whether child has ever attended school (self-reported), whether child missed school in last schooling week for a reason other than public holiday, school closure or teacher absence (self-reported), and whether child was recorded as attending more than 80% of open school days in May 2012 school register (endline for primary schools only). Romano-Wolf adjusted p-values are shown in brackets and standard errors, clustered by village, are shown in parentheses. All regressions control for the baseline value of each imbalanced variable in Table 2, a dummy to indicate this variable was missing and interactions of each with period. Note that asterisks refer to standard (non-adjusted) p-values.

*** significant at the 1% level

** significant at the 5% level

* significant at the 10% level.

Table A2: Replication of Main Effects with No Controls

	Ever Attended (age 6-21) (1)	Missed Last Week (age 6-21) (2)	School Attendance (age 6-21) (3)	Primary Completed (age 14-21) (4)	First Grade Sec. Completed (age 14-21) (5)
Treatment X Midline	0.04***	-0.04		0.04	-0.03
	(0.02)	(0.03)		(0.04)	(0.04)
	[0.00]	[0.32]		[0.39]	[0.39]
Treatment X Endline	0.05**	-0.04	0.05	0.06	-0.07
	(0.02)	(0.04)	(0.04)	(0.05)	(0.05)
	[0.00]	[0.39]	[0.27]	[0.32]	[0.27]
Midline	0.05***	0.05***		0.26***	0.21***
	(0.01)	(0.02)		(0.04)	(0.03)
Endline	0.09***	0.06***		0.39***	0.32***
	(0.01)	(0.02)		(0.04)	(0.04)
Observations	8963	7670	1943	3284	2992
Baseline mean	0.82	0.44		0.50	0.11

Source: Authors' calculations based on 2009, 2011, and 2012 household survey data.
Note: Table presents estimates of effect of CCT at midline and endline. Each column is a separate regression. The analytic sample is children who were 6 to 21 or 14 to 21 at the time of survey. Dependent variables are a dummy to indicate whether child has ever attended school (self-reported), whether child missed school in last schooling week for a reason other than public holiday, school closure or teacher absence (self-reported), whether child was recorded as attending more than 80% of open school days in May 2012 school register (endline for primary schools only), whether child completed primary (self-reported), and whether child completed first grade of secondary (self-reported). Romano-Wolf adjusted p-values are shown in brackets and standard errors, clustered by village, are shown in parentheses. Note that asterisks refer to standard (non-adjusted) p-values.

*** significant at the 1% level

** significant at the 5% level

* significant at the 10% level.

Table A3: Replication of Main Effects with Alternative Age Group for School Completion Variables

	Primary Completed (age 12-21) (1)	First Grade Sec. Completed (age 12-21) (2)
Treatment X Midline	-0.00	-0.05
	(0.03)	(0.04)
	[0.95]	[0.30]
Treatment X Endline	0.01	-0.09*
	(0.03)	(0.05)
	[0.81]	[0.07]
Midline	0.38***	0.18***
	(0.04)	(0.03)
Endline	0.55***	0.26***
	(0.04)	(0.05)
Observations	4409	4120
Baseline mean	0.36	0.08

Source: Authors' calculations based on 2009, 2011, and 2012 household survey data.

Note: Table presents estimates of effect of CCT at midline and endline. Analysis replicates Table 4 exactly for all outcomes except changing the age group presented. Each column is a separate regression. The analytic sample is children who were 12 to 21 at the time of survey. Dependent variables are a dummy to indicate whether child completed primary (self-reported), and whether child completed first grade of secondary (self-reported). Romano-Wolf adjusted p-values are shown in brackets and standard errors, clustered by village, are shown in parentheses. All regressions control for the baseline value of each imbalanced variable in Table 2, a dummy to indicate this variable was missing and interactions of each with period. Note that asterisks refer to standard (non-adjusted) p-values.

*** significant at the 1% level

** significant at the 5% level

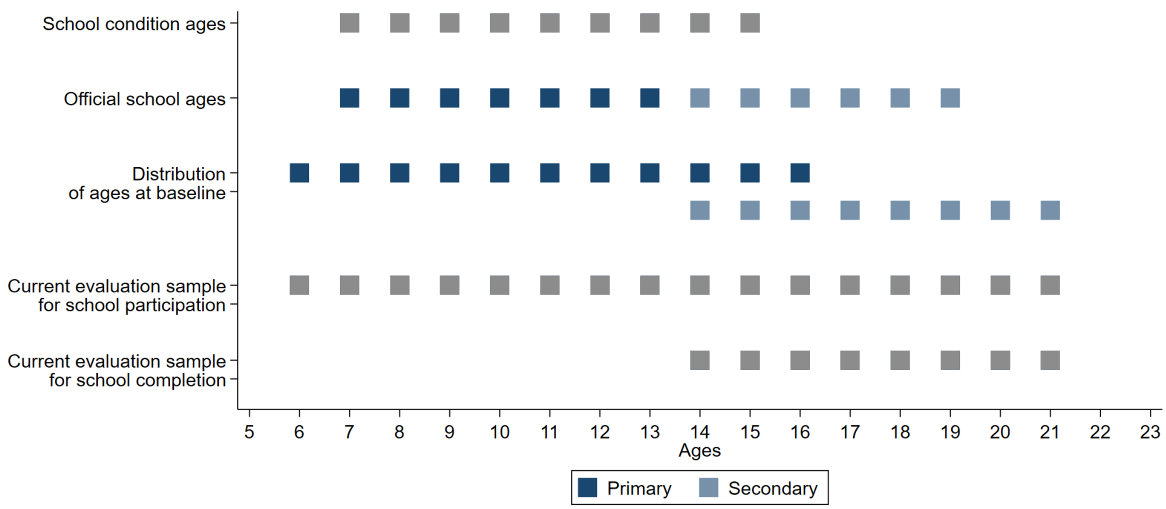
* significant at the 10% level.

Table A4: Correlation Matrix for Dimensions of Vulnerability

	Female	Poorest	Not passed
Female	1.000		
Poorest	0.048	1.000	
Not passed	0.010	0.096	1.000

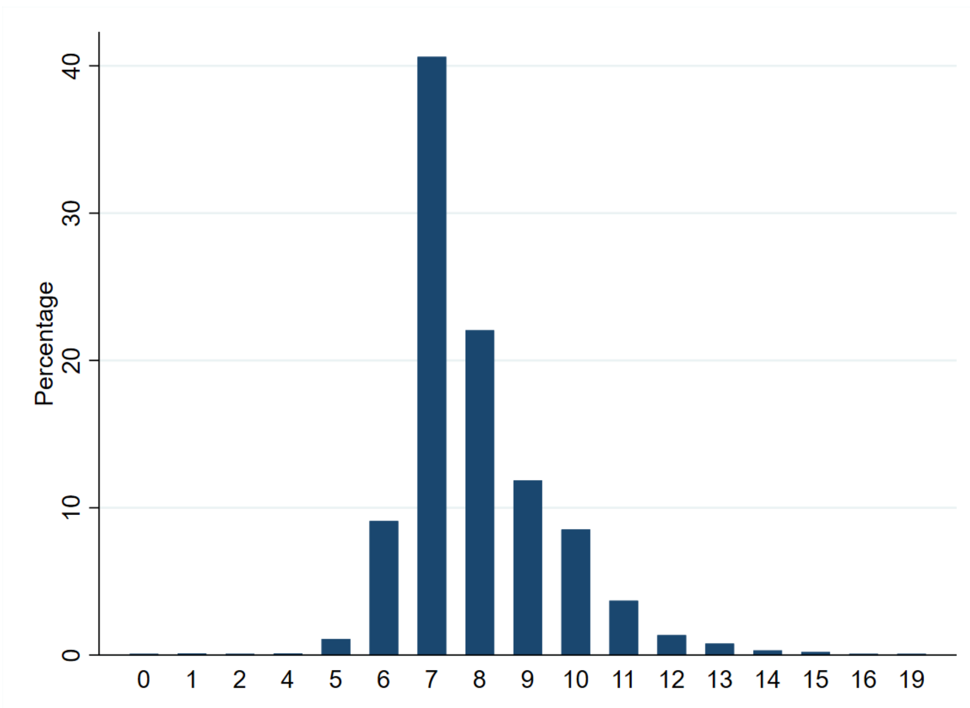
Note: Table presents pairwise correlations for dimensions of vulnerability at baseline. Poorest indicates child is from a household with below-median consumption. “Not passed” indicates child did not pass a primary exam.

Figure A1: Relevant Age Ranges for This Evaluation



Note: This figure shows ages for which (1) the CCT schooling condition applies, (2) the official ages for primary and secondary school in Tanzania, (3) the actual distribution of ages at baseline in our data, (4) the ages we include in our sample for school participation and attendance outcomes, and (5) the ages we include in our sample for school completion outcomes.

Figure A2: School Start Ages



Note: Plot shows distribution of ages at school start for 6-21 age group at baseline.

Figure A3: Density of Absent Days by Treatment Group

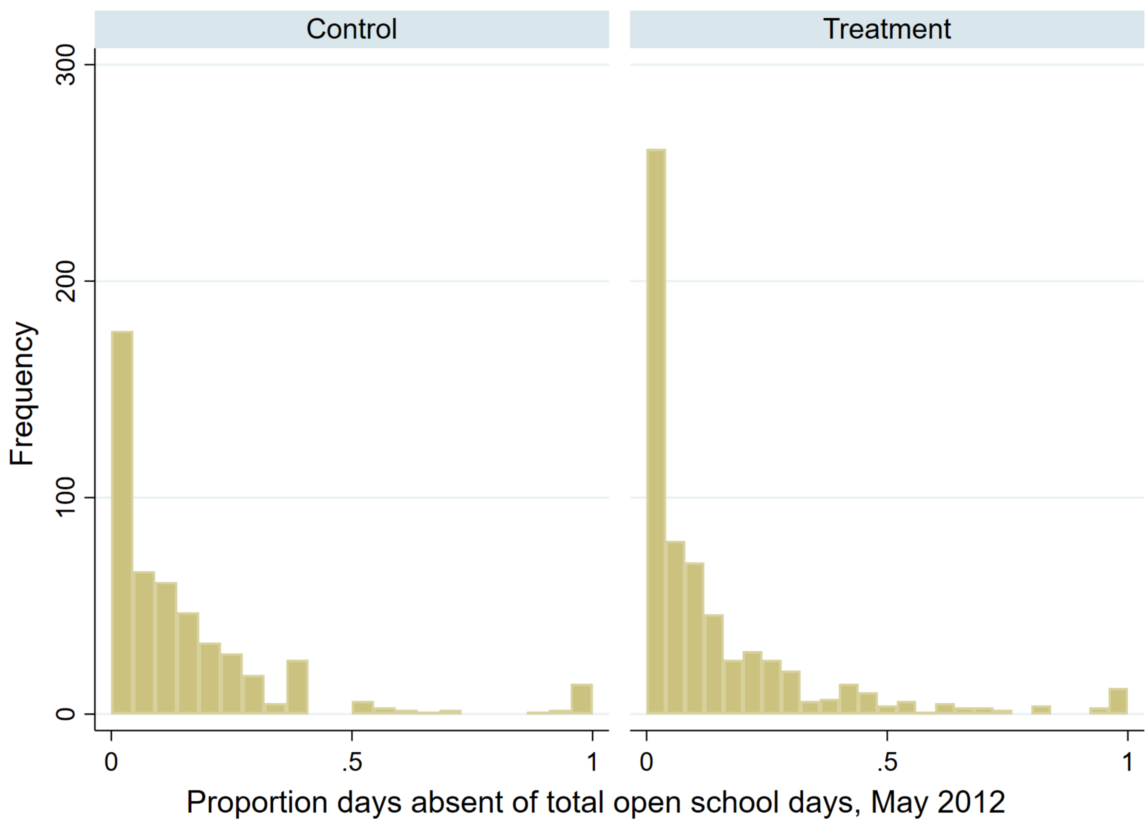
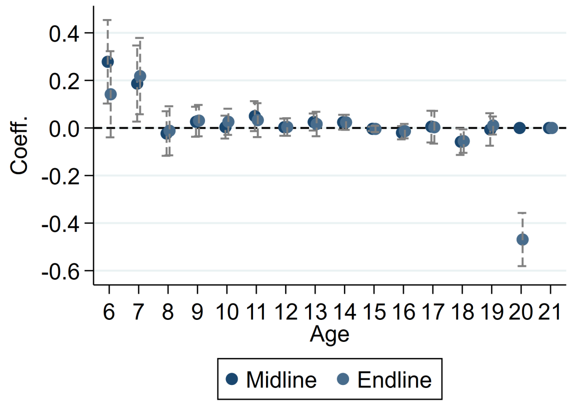
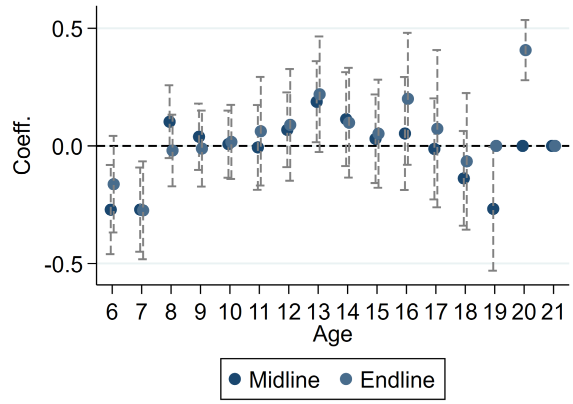


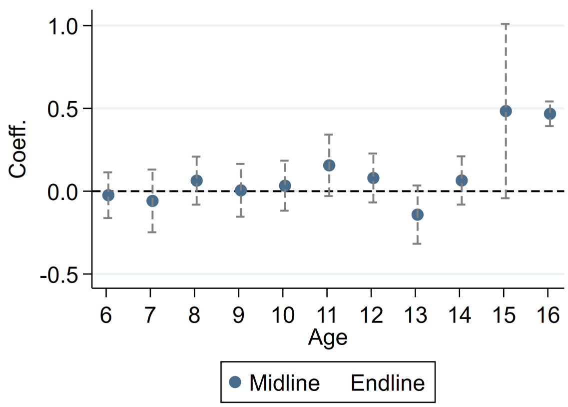
Figure A4: Treatment Effects by Baseline Age



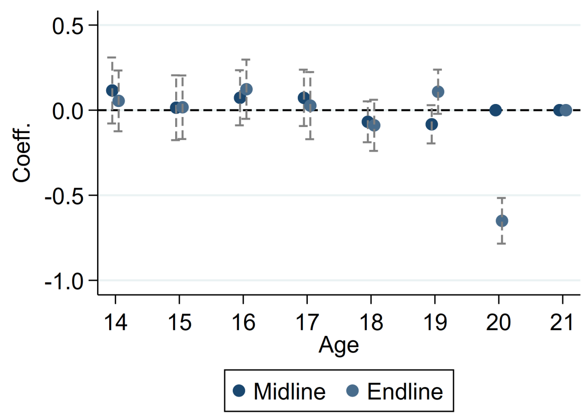
(a) Ever Attended



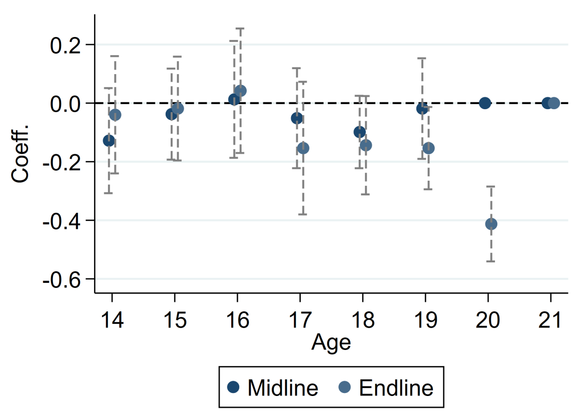
(b) Missed Last Week



(c) School Attendance



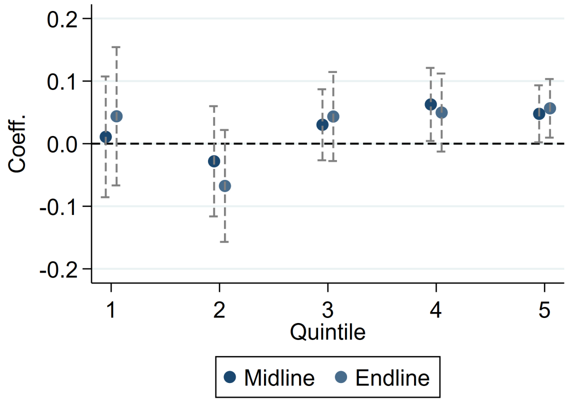
(d) Primary Completed



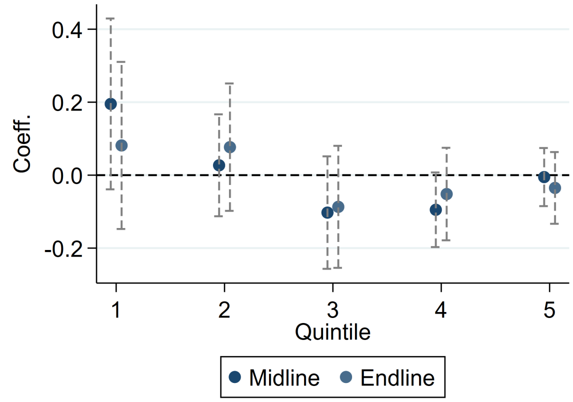
(e) Attended Secondary

Note: Plots show treatment effects and 95% confidence intervals for each outcome by baseline age. For ever attended, missed last week and school attendance the omitted category is 6 year olds in comparison villages; for primary complete and attend secondary the omitted category is 14 year olds in comparison villages.

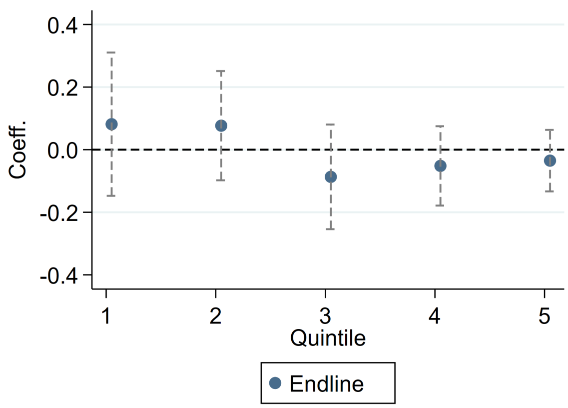
Figure A5: Treatment Effects by Household Consumption Quintile



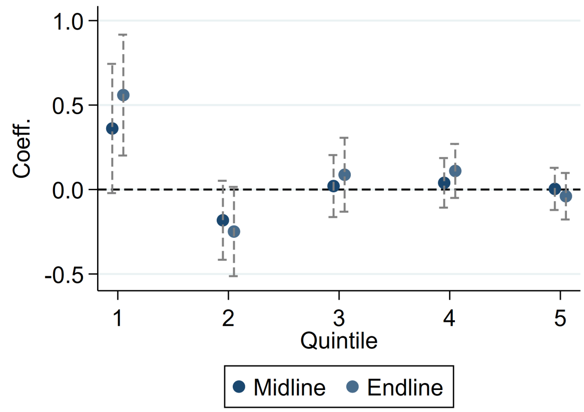
(a) Ever Attended



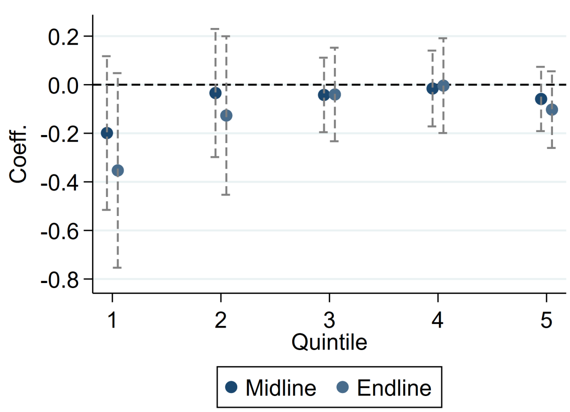
(b) Missed Last Week



(c) School Attendance



(d) Primary Completed



(e) Attended Secondary

Note: Plots show treatment effects and 95% confidence intervals for each outcome by household poverty status. Poverty status calculated by grouping baseline household consumption into quintiles. In each regression the omitted category is quintile 1 in comparison villages.

A.1 Data Processing Notes

For three of our outcomes, an affirmative response in one period implies the same response in a subsequent period while a value of zero in a latter period implies zero for an earlier period. For example, if an individual reports having ever attended school in period 1, they must also have ever attended at periods 2 and 3, regardless of whether they attrit from the sample. Similarly if an individual reports never having attended school in period 3, they must also never have attended at periods 1 and 2. This is true for the outcomes ever attended, completed primary and completed first grade of secondary. To ensure responses are logically consistent in our data, we use an "imputation"/carry forward strategy for these three outcomes.

For each outcome we replace with a value of 1 all period 2 individuals that have a value of 1 in period 1, then we replace with a value of 1 all period 3 individuals that have a value of 1 in period 2 or period 1. Further, we replace with a value of zero all period 1 individuals that are missing data but have a value of zero in periods 2 or 3, and we replace with a value of zero all period 2 individuals that are missing data but have a value of zero in period 3. This means that we have reshaped the data so that if an individual attrits from the sample but had a value of 1 on one of these outcomes in a previous period, they now have an observation for the subsequent period. If after carrying this out, an individual is still missing all 3 outcomes in period 2 or period 3, that individual is considered to have attrit.

For some individuals, this process also requires imputing an age for that period. If an individual is missing an age for period 2,[21] we impute an age value that is 2 greater than their baseline age. If an individual is missing an age for period 3, we impute an age value that is 3 greater than their baseline age. Lastly, since all other relevant covariates are measured at baseline and are thus time-invariant, we carry forward those values if they are missing. The following table presents the percentage of observations that are carried forward in this way for each age sample:

Table A5: Data Processing

Ever Attended	Primary Completed	First Grade Sec. Completed	Ages
Ages 6-21			
15.2	9.1	5.8	14.4
Ages 14-21			
24.1	12.0	4.5	23.4

[21] Date of birth information is only available for children who are 7 years or younger.

A.2 Causal Forest Procedure

To complement our primary approach to analyzing heterogeneity, we use a causal forest machine learning algorithm (Athey and Imbens, 2016; Wager and Athey, 2018). In contrast to our traditional ad hoc approach of picking characteristics and testing them separately as treatment by covariate interactions, machine learning allows for higher-dimensional prediction using a larger set of information on sample participants. In our specific case, we have two primary goals; first, assessing the relative importance among a large set of covariates in predicting treatment effects, and second, identifying a sub-group based on observable characteristics for whom the treatment was particularly effective.

In adopting the approach of Athey and Imbens (2016) and Wager and Athey (2018), we describe this method in brief and refer readers to those papers for further details. The basic approach is to divide the endline sample into a training sub-sample used to grow the causal forest predictions, and a testing sub-sample to which predictions are applied. The causal forest procedure begins by drawing a random sub-sample of the training set and dividing it into bins based on values of covariates. As an example the algorithm might start with gender, dividing the sample into males and females, and then further into total household consumption (a measure of household poverty), resulting in a "tree" where branches are the bins and nodes are the points at which the branch further subdivides. The algorithm divides each training sub-sample into bins in a way that maximizes variance in treatment effects subject to a within-bin variance penalty, choosing random splits through which to further subset the data. Specifically, in growing the tree it aims to maximize an objective function which Athey and Imbens (2016) show can be used to approximate the expected mean squared error in the model; thus, the algorithm predicts treatment effects by deciding which conditioning covariates are important in terms of their goodness-of-fit. The algorithm separates the training set into two sub-samples in order to reduce the chance of mistaking random variation for signal in what is termed an "honest" approach. This iterative process is repeated across tens or hundreds of thousands of sub-samples (the "forest"), each time resulting in a predicted treatment effect for observations that take the value of covariates used to make the bins.

We use the R package grf (Tibshirani et al., 2020) to estimate heterogeneous treatment effects at endline based on a large number of observable characteristics. For each outcome, we grow a forest of 100,000 trees, set the minimum number of observations in each leaf at the default value of 5, and split the training sample into equal-sized sub-samples. We include as conditioning covariates all characteristics presented in Tables 2 and 3, while also including additional characteristics that have been analyzed in previous studies of the effects of conditional cash transfers on education outcomes; these are indicator variables for foster status, orphan status, and distance to school.