When Your Bootstraps Are Not Enough: How Demand and Supply Interact to Generate Learning in Settings of Extreme Poverty

Alex Eble and Maya Escueta*

April 2024

Abstract

In settings of extreme poverty, how do demand and supply combine to produce child learning? In rural Gambia, caregivers with high aspirations for their children's future, measured before children start school, invest substantially more than others in children's education. Despite this, essentially no children are literate or numerate three years later. When villages receive a highly impactful, teacher-focused supply-side intervention, however, children of high-aspirations caregivers are 25 percent more likely to achieve literacy and numeracy than others in the same village. We estimate patterns of substitutability and complementarity between demand and supply in generating learning that change with skill difficulty.

*Eble: Teachers College, Columbia University, 525 W 120th, New York, NY 10027, and National Bureau of Economic Research; email: eble@tc.columbia.edu. Escueta: American Institutes for Research, 100 Europa Drive Suite 315, Chapel Hill, NC 27517; email: mescueta@air.org. We are grateful to the collaborators on the SCORE Trial - Chris Frost, Alpha Camara, Baboucarr Bouy, Momodou Bah, Maitri Sivaraman, Pei-Tseng Jenny Hsieh, Chitra Jayanty, Tony Brady, Piotr Gawron, Stijn Vansteelandt, Peter Boone, and Diana Elbourne - as well as our research team in the Gambia, particularly Lamin Janneh; see Boone et al. (2015) and Eble et al. (2021). This paper reports analysis of data collection which was approved by the Institutional Review Board of the London School of Hygiene and Tropical Medicine, protocol number 8767. We received helpful input from Kimi Chan, Beezer Cheble, Sarah Cohodes, Esther Gehrke, Joshua Goodman, Naureen Karachiwalla, Adrienne Lucas, Josh Merfeld, Kate Orkin, Phillip Ross, Felipe Valencia, and seminar audiences at the Columbia Committee on the Economics of Education, the Columbia Population Research Center, the Gui2de/World Bank joint seminar, the NBER Program on Children, the New York Nanodev conference, Northeastern University, Princeton University, the RISE 2021 Annual Conference, and the University of Chicago Workshop on Education. Authors have no conflicts of interest to declare. This paper updates and replaces the previously circulated paper "Aspirations, Education, and Extreme Poverty." Eble acknowledges support from the National Academy of Education (NAEd) and the NAEd/Spencer Postdoctoral Fellowship Program. The authors have no conflicts of interest to declare. Keywords: human capital; education; poverty; complementarity; demand; supply; learning. JEL codes: I25, O15, I28, I24, O12.

1 Introduction

Many families wish to provide better lives for their children than experienced by those in previous generations. Education is a primary lever families use to achieve this goal. Intergenerational educational mobility is an important source of economic mobility, both in Europe and North America (Black et al., 2011; Chetty et al., 2014, 2017) and, as shown more recently, in many low- and middle-income countries (Azam and Bhatt, 2015; Asher et al., 2018; Alesina et al., 2021). There is an established empirical link between family demand for education and child learning in some of these countries (c.f. Foster and Rosenzweig 1996; Behrman 2010; Jensen 2010; Beaman et al. 2012). It is not clear, however, whether this relationship holds in settings – particularly those characterized by extreme poverty – where complementary supply-side inputs are often absent or of extremely low quality.

In this paper, we investigate when and how family demand matters for child learning in a very low-income setting. The analysis proceeds in two steps. First, we ask: if a caregiver in such a setting wants to raise their children's learning levels, how much learning can they bring about on their own? We answer this question by estimating the observational relationship between measures of latent demand at baseline and subsequent educational investment and skill acquisition. Second, we ask: does the relationship between family demand and child learning change when the constraint of very low-quality educational supply, common in such settings, is relaxed? We answer this question by generating a causal estimate of how the demand–learning relationship changes in the face of a randomly assigned shock to the quality of educational supply. Our measures of learning focus on reading and math skills that are "developmentally meaningful;" in other words, lower-level skills which influence the child's ability to acquire higher-level skills and succeed in later years of schooling (Duncan et al., 2007; Muralidharan et al., 2019; Nelson III and Gabard-Durnam, 2020). In this analysis, we also explore patterns of complementarity and substitutability between demand and supply at different levels of skill acquisition to better understand how demand and supply interact to generate learning.

The data we use follow children and their caregivers in rural Gambia during a crucial period of child development in terms of skill acquisition – the three year period beginning immediately before primary school. The data come from a census of families in 169 villages in the two central regions of The Gambia. We track families who, at the time of this census, intended to enroll at least one of their children in the first grade for the first time in the upcoming fall semester (2015). Over the next three years, data were collected on the child's school enrollment and school-related time use, as well as on the family's educational expenditure for the child. At endline (spring 2018), children took reading and math tests that are highly sensitive to measuring early and intermediate skill acquisition. The tests include several skills that either precede or comprise literacy and numeracy and that are necessary for the acquisition of higher-level skills such as abstract reasoning and composition (Werker and Tees, 2005; Duncan et al., 2007; Nelson III and Gabard-Durnam, 2020).[1]

We use two measures of family demand, collected at baseline: caregivers' desire for their child's ultimate educational attainment and caregivers' aspirations for the child's future career. These measures draw on a series of theoretical and empirical studies showing that such aspirations – a specific type of desire for the future – are strongly linked to greater investment in children and higher educational outcomes (cf. Beaman et al. 2012; Bernard et al. 2014; Genicot and Ray 2017; Lybbert and Wydick 2018; La Ferrara 2019).[2] In our study, these serve as coarse measures of an important but latent variable: family demand for helping the child achieve a better life than experienced by previous generations.

We find that families who express high demand at baseline go on to invest more in their child's education over the following three years. Relative to other families, these families are three to six percentage points more likely to enroll their children in school in the first two years of the study. In the final year of the study, when essentially all children are enrolled in school, high-demand families spend significantly more money on their children's education than other families, and

[1]The two tests are Early Grade Reading and Math Assessment-style tests, also known as EGRA and EGMA tests. See Platas et al. (2014) and Dubeck and Gove (2015) for details on their development, implementation, and limitations.
[2]Fruttero et al. (2021) summarizes recent empirical research on this topic.

their children spend more time each day on school-related tasks.

At the end of three years, the children of these high-demand families score 0.28-0.30 standard deviations (SD) better on tests of reading and math ability. The SD metric is commonly used to measure learning gains, particularly in the many hundreds of impact evaluations of educational interventions conducted in low-income settings (McEwan, 2014; Glewwe and Muralidharan, 2016; Ganimian and Murnane, 2016; Evans and Yuan, 2022). Seen through the lens of these studies, our estimates suggest substantially higher learning levels among children of high aspirations caregivers.

Sadly, we show that the true mapping from family demand to endline learning in rural Gambia is close to zero. Using measures of skill acquisition to characterize the demand–learning relationship rather than the SD metric, we estimate a precise zero for the relationship between family demand and endline levels of literacy and numeracy. We estimate very small positive associations for the acquisition of other developmentally meaningful pre-literacy and pre-numeracy skills such as reading short words or calculating basic sums. In short: after three years of schooling, essentially none of the children in these areas of The Gambia possess any of the skills necessary for literacy and numeracy – and expected of grade 2 students in Gambian schools. Our findings therefore belie the notion that families in such settings merely need to wish and try harder to "pull themselves up by their bootstraps" to realize their desires for their children's futures.

This surprising result also illustrates the difficulty of using the SD metric to assess learning gains in settings where baseline learning levels are close to zero. In such settings, even a very small absolute change in test scores translates into a large relative gain. This increases the risk of erroneously positive conclusions about the importance of a given input.[3]

Our estimates are also a likely upper bound on the status-quo relationship between demand for education, as measured by aspirations, and learning outcomes. This is because the most plausible

[3]Multiple investigations have have documented problems with the SD metric; see, for example, Singh (2015; available at https://blogs.worldbank.org/impactevaluations/how-standard-standard-deviation-cautionary-note-using-sds-compare-across-impact-evaluations, accessed June 2, 2021), Filmer et al. (2020), and Evans and Yuan (2022). In addition, research on U.S. schools has found an inverse relationship between the learning represented by a given effect size estimate and the baseline learning level in the focal population (Hill et al., 2008).

potential unobservable confounders – for example, unobserved wealth or family preferences – are likely to be positively correlated with both the aspirations we study and educational outcomes (Bernard et al., 2014; Ross, 2019). Were such traits to influence our estimates, the true relationship would be even smaller than what we measure.

In the second part of our paper, we show how demand and supply interact to generate learning. A highly-resourced, teacher-focused supply-side educational intervention was randomly assigned to half of study villages.[4] We use this as an exogenous shock to the quality of education supplied in these villages. In the presence of high-quality educational supply, we show that the mapping from baseline family demand to subsequent child learning is large and developmentally meaningful.

In villages benefitting from the intervention, children of families with high educational aspirations at baseline are 25 percent more likely to achieve literacy and numeracy than other children *in the same village*. This result is robust to alternative measures of literacy and numeracy, and also holds for these children's acquisition of related skills, such as the number of words the child can correctly read per minute. In other words, we show that demand does matter for learning, even in this extremely low-income setting, when educational supply is of high quality. On the other hand, we find that children of families with high career aspirations at baseline do not have significantly different learning outcomes than other children in the same village, conditional on the village receiving the intervention.

The final part of our paper studies the technology of early skill formation in such very low-income settings. To do so, we exploit the fact that the tests we use to measure learning capture rich, granular data on each child's formation and mastery of a sequence of skills. These skills increase in difficulty over the course of the test and characterize a child's progress on the path towards achieving foundational literacy and numeracy (Platas et al., 2014; Dubeck and Gove, 2015). We use these data to show how demand and supply combine to form skills at varying levels of difficulty.

We find patterns of complementarity and substitutability between demand and supply in producing these skills which change as skill difficulty increases. For the lowest-level reading and math

[4] Eble et al. (2021) show that the intervention yielded transformative learning gains for all students in these villages.

skills, we find substitutability between aspirations (both educational and career) and educational supply. For higher level skills, we find evidence of complementarity between educational aspirations and educational supply. These results show that, in such settings, a large improvement in the quality of educational supply moves out the frontier from which families with high demand for education invest in their children, shifting the relationship between family demand and the child's acquisition of foundational literacy and numeracy skills.

We address two potential alternative explanations for the results in the second part of our paper: that the aspirations measures might merely capture unobserved child ability or household wealth, both of which might be associated with greater learning when the quality of educational supply increases. Unlike in the first part of the paper, we cannot use a bounding argument, because the intervention could either substitute for or reinforce the role of any unobserved factors. Instead, we explore the likely magnitude of these contributions. Several facts make it exceedingly unlikely that caregiver aspirations are merely a proxy for child ability, most notably: an extremely low proportion of caregivers had ever gone to school or were able to read; aspirations were measured prior to the child starting school; and even after children begin attending school, caregivers in such settings often have highly inaccurate beliefs about child ability (Dizon-Ross, 2019). We also present a series of empirical analyses documenting that our aspirations measures are only mildly correlated with our observed measures of wealth, and are unlikely to capture unobserved family wealth that interacts with educational supply. This is further corroborated by the extremely – and largely uniform – low levels of wealth in these areas.

Our study shows how the inputs of families (through the aspirations and investments of caregivers) and school systems (through the availability of quality educational inputs) combine to create foundational literacy and numeracy skills at a crucial juncture in children's lives. Our first contribution is to characterize the status-quo relationship between family demand, family inputs, and learning in a very low-income setting. Many families in these areas dedicate substantial amounts of household resources, both money and time, towards helping their children learn. Despite these investments, students in the status quo areas are highly unlikely to master skills crucial for their developmental trajectory – specifically literacy, numeracy, and related skills – in this pivotal period of cognitive development. This result is extremely troubling because children who fail to master these skills in this period of their lives face a very low probable ceiling on their learning trajectory.[5]

Our second contribution is to show that family demand *can* map onto meaningfully greater learning, even in very low-income settings, in the presence of adequate supply. The dramatic change in the quality of supply generated by the intervention provides multiple necessary inputs that are absent in the status quo. These inputs shift the impact of high-aspirations families' investments in their children, moving these children from a status quo state of having somewhat greater likelihood of mastering very rudimentary skills, relative to other children in their village, to having a substantially greater likelihood of mastering higher-level reading and math skills, including literacy and numeracy.

Our results add to knowledge, mostly from studies in higher-income contexts, on how the effects of impactful educational interventions vary with baseline family characteristics. Krueger and Whitmore (2001) show that small class sizes have a larger benefit for children of lower-income families, and Jackson et al. (2016) show similar heterogeneity in the effects of increases in school spending. Walters (2018) and Cohodes and Parham (2021) show the effects of access to charter schools also vary by race and socioeconomic status. We show that family traits can also determine how much supply-side interventions affect educational outcomes in a much lower-income context.

The study also advances general understanding of how the demand-side and supply-side interact to generate learning in low-income settings (cf. Jensen 2010; Glewwe and Muralidharan 2016; Muralidharan et al. 2019; Romero et al. 2020). Building on recent studies of complementarities between educational inputs on the supply side in similar settings (Mbiti et al., 2019; Kerwin and

[5]From a child development perspective, the age range we study (ages 6-11) is critical for children's learning and cognitive development; children who do not acquire foundational reading and math skills during this period have a very difficult time acquiring these skills later in life (Knudsen, 2004; Werker and Tees, 2005; Nelson III and Gabard-Durnam, 2020). These skills, in turn, play an important role in the child's ability to acquire higher-level skills and succeed in the later years of school (Duncan et al., 2007; Wolf and McCoy, 2019). Further, government teachers are incentivized to teach at grade level, rather than address gaps in skills meant to be taught in earlier grades (Banerjee et al., 2017a; Muralidharan et al., 2019). Children who reach upper grade levels without mastering these skills – in the Gambian case, essentially all children in the status quo – are therefore highly unlikely to ever acquire these skills (Cunha and Heckman, 2007; Pritchett and Beatty, 2015; Muralidharan et al., 2019; Niaz Asadullah et al., 2019).

Thornton, 2021), our analysis reveals patterns of substitutability and complementarity between demand and supply in children's acquisition of reading and math skills.

Finally, these results also contribute to the growing body of work on the role of aspirations in education and development (cf. Dalton et al. 2016; Genicot and Ray 2017; Lybbert and Wydick 2018; Fruttero et al. 2021; Serneels and Dercon 2021). In a process known as "aspirations failure" or "aspirations frustration" (Genicot and Ray, 2017; Ross, 2019; McKenzie et al., 2022), the link connecting aspirations to investment and outcomes can fracture when the hoped-for outcome is so distant as to seem futile, which in turn depresses related investment. We show that in rural parts of The Gambia, an analog "system failure" can occur when, despite a robust mapping from aspirations to investment, the mapping from investment to key developmental outcomes collapses in the absence of other necessary inputs.

2 Setting and data

In this section, we describe the setting of our study, the data we analyze, and our measures of learning and aspirations.

2.1 Setting

Our study takes place in small, rural settlements in the Lower River and North Bank regions of The Gambia. The Gambia is located in West Africa, with Senegal on its border to the north, east, and south, and the Atlantic Ocean to its west.[6] Its population is roughly two million people, and its geographic area covers roughly 11,300 square kilometers (CIA, 2019). It is a former British colony and served as a major hub for the trans-Atlantic slave trade (Wright, 2015). The devastation and historical impacts of this legacy are important contributors to the fact that The Gambia is very income poor, with per-capita GDP estimated to be $716 in 2018. The country's main sources of economic activity are currently agriculture, tourism, remittances, and foreign aid.

In addition to income poverty, the country's education levels are also very low. In 2013, the Demographic and Health Surveys estimated that only 26.7 percent of adults living in rural areas

[6] In Figure A.1, Panel A, we show a map of The Gambia's location on the African continent.

were literate, and roughly half of adults in these areas had never been to school (The Gambia Bureau of Statistics and ICF International, 2014). Other national assessments of children's reading and math abilities have shown that learning levels among children in The Gambia are dramatically lower than in other countries in the region (Sprenger-Charolles, 2008).

The population of our study comes from a census of all villages in these two regions meeting a series of pre-specified eligibility criteria. We began with the universe of villages in these two regions which had between 10 and 300 households according to the 2013 national census.[7] Of these villages, we enrolled those which had at least 10 eligible children resident in the village at the time of enumeration in early 2015.[8] Children were eligible if, at time of enumeration, they were between the ages of 6 and 8, they had not yet entered the first grade, and their primary caregiver intended to enroll them in the first grade in the coming academic year. Ultimately, 169 villages across the two regions were enrolled in the trial. The participants in our study were all children in the village meeting these eligibility criteria, and each eligible child's primary caregiver.

Because presence in this sample is conditional on the caregiver intending to enroll the child in school in the coming year, the educational trajectory of participants may differ from the population in these areas. When abstracting from our sample to the broader population of children in these areas, we make the following assumption: the trajectory of literacy and numeracy skills among excluded children is unlikely to be dramatically better than of study participants, though it could be either similar, or worse. This stems from the fact that excluded children will enter school later than study children, and later school entry corresponds to worse academic outcomes in similar settings (Glewwe and Jacoby, 1995; Bommier and Lambert, 2000).

There were 4,518 children enumerated at baseline, 3,825 for whom we have endline test scores. Because our focus is on child learning over the course of the study, these 3,825 children comprise our study population.[9] In the next section, we describe the characteristics of these children and

[7] In Figure A.1, Panel B, we show a map of The Gambia indicating the regions in which these villages are located.

[8] There were 323 total villages to begin with. Of these, 113 had too few children to be eligible. The study excluded a further 41 of the remaining villages to create buffer zones between villages in order to ensure no potential for spillover between villages, i.e., caregivers of children in control villages instructing their children to walk into an intervention village and avail themselves of the intervention there.

[9] Baseline aspirations do not predict attrition at the endline test.

their families.

2.2 Intervention

Clusters of villages were assigned to be in either the intervention or control group. Randomization was stratified by region (Lower River and North Bank) and distance to main road in each region (above or below median). Those in the intervention arm received a highly-resourced intervention providing an after-school, remedial education program delivered by para teachers. This program began in early 2016 and continued until the first week of May 2018. The program bundled together multiple teacher-focused prongs known to work in isolation. It began by hiring para teachers, either from within the village or nearby (Kingdon and Sipahimalani-Rao, 2010; Muralidharan and Sundararaman, 2013). It trained them to use scripted lessons (Piper et al., 2014; Banerjee et al., 2017a) to deliver after-school, supplementary education for 12 hours per week over the course of the study, following the official Gambian curriculum as children progressed through school. These para teachers were regularly monitored with a focus on "coaching," that is, improving their instructional capacity and ensuring student learning (Kraft et al., 2018; Piper et al., 2018). Eble et al. (2021) show that this intervention was highly effective at raising learning levels for all children in villages randomly assigned to receive it.

2.3 Data

Data were collected from participants (children and their caregivers) over the period from January 2015 to June 2018. Participants were enumerated in early 2015 and randomization occurred in late 2015. In Table 1, we present a few key demographic characteristics of the children in our sample, overall and separately by the arm of the trial into which they were randomized. We refer to children enumerated in villages that were subsequently randomized to not receive the intervention (i.e., the control group) as the "status quo" group. We refer to children enumerated in villages subsequently randomized to receive the intervention as the "intervention" group. At baseline,

Table 1: Demographic characteristics

	(1) All	(2) Status quo	(3) Intervention
Child is female	0.50	0.51	0.48
Caregiver can read simple sentence	0.08	0.08	0.08
Caregiver is not child's mother	0.23	0.22	0.23
Books found in house	0.67	0.65	0.69
Caregiver education			
Never been to formal schooling	0.76	0.77	0.76
At least some primary education	0.16	0.15	0.16
At least some junior secondary education	0.06	0.06	0.06
At least some senior education, or more	0.02	0.02	0.02
Household wealth			
House is made of all natural materials	0.06	0.05	0.08
House is made of partially synthetic materials	0.68	0.68	0.68
House is made of all synthetic materials	0.26	0.28	0.24
Observations	3,825	2,045	1,780
Joint F-statistic		0.652	
(p-value)		(p= 0.688)	

Table 1 notes: this table presents select demographic characteristics for children in our sample, both overall (column 1) and then separately by the treatment status to which they were randomized (columns 2 and 3, respectively). The joint F-statistic is a test of the null that these variables together are not jointly predictive of the child's randomization status to the intervention (treatment) or status quo (control) group, clustering errors by clusters of contiguous villages. All variables in this table, except for the number of observations, are binary, with 0 = No and 1 = Yes.

fewer than 25 percent of primary caregivers in either group had ever been to school.[10] This is lower than average levels in The Gambia (The Gambia Bureau of Statistics and ICF International, 2014), consistent with the fact that the areas in which the study took place have lower income levels, are more remote, and are less well-served by the government than many others in the country. We observe a simple proxy for wealth: whether the floor, walls, and roof of the home are made of synthetic materials (also used in Fazzio et al., 2021), with roughly one quarter of households living in homes constructed entirely out of synthetic materials. There is balance between randomization groups in these and other observable characteristics, as shown in the p-value of the joint F-test that these characteristics predict group membership, reported at the bottom of the table (Bruhn and

[10]We focus on caregivers, as opposed to parents, because early fieldwork suggested that the most important person for the child's development is the primary person from whom the child receives their day-to-day care. This is often, but not always, the parent. In our data, roughly 75% of caregivers are mothers, 11% are grandmothers, and the rest are various other members of the household in which the child lives.

McKenzie, 2009).

We collect three types of data on family investment in the the child's schooling over the course of the study. We use these both to document family investment behavior in the child, and to show the performance of caregiver aspirations as a measure of the family's latent demand for helping the child on to a better future. If these were good measures of demand, we would expect them to positively correlate with subsequent investment behavior in the child's education.

Our measures capture investment on both the extensive and intensive margins. The first is on the extensive margin: it captures the child's enrollment in school in a given year, and is collected at the end of each academic year. The second and third measures are on the intensive margin. These capture the caregiver's annual financial expenditure on the child's education (comprising teacher "top-up" fees, school materials such as stationery, and other related costs), and the proportion of the child's waking hours on an average weekday spent on school-related tasks, which we refer to as "time use." These were both collected at the end of the third year.

2.4 Measuring learning

We measure child learning at endline with tests conducted in May and June of 2018. These tests were EGRA- and EGMA-style assessments – short for Early Grade Reading and Math Assessments, respectively – administered to each study child one-on-one as per test guidelines (Platas et al., 2014; Dubeck and Gove, 2015). They are designed to precisely measure the acquisition of a series of early grade reading and math skills which are either precursors to or components of achieving literacy and numeracy. They are also highly sensitive to capturing learning at the earliest stages, minimizing the risk of floor effects in this type of assessment. Floor effects are a particularly salient risk when learning levels are very low, as is common in settings like ours.

To capture these data, each test is comprised of questions that belong to six different "subtasks," each of which measures a child's mastery of one such skill. In Table 2 we list each subtask and describe the skill it evaluates. As the number of the subtask rises, so does the difficulty of the skill it measures. For example, reading subtask 1 focuses on letter sound identification, a precursor to

Table 2: Test subtasks

Reading		*Math*	
Subtask	**Evaluates the child's ability to**	**Subtask**	**Evaluates the child's ability to**
1	Read a letter's sound (e.g., "eh" for e)	1	Read a number (e.g., 1, 5, 22)
2	Differentiate sounds (e.g., which word starts with a different sound: book, dog, or boy)	2	Choose the larger number (e.g., 7 or 5)
3	Read a made-up word (e.g., tob)	3	Complete a sequence (e.g., 2 4 6 __)
4	Read a familiar word (e.g., but)	4a	Do simple addition (e.g., 3+2)
		4b	Do two- and three-digit addition (e.g., 38+26)
5a	Read a short passage	5a	Do simple subtraction (e.g., 5-3)
5b	Answer questions on the passage's content	5b	Do two- and three-digit subtraction (e.g., 59-37)
6	Listen to a different short passage, answer questions on the passage's content	6	Solve a simple word problem read aloud

Table 2 notes: this table describes the individual "subtasks" within the reading (EGRA) and math (EGMA) tests administered at endline. The full test papers are given in Appendix A; the relevant subtask number for each block of questions is indicated in the test papers.

(and easier than) the skill evaluated in reading subtask 4, familiar word recognition. We provide the full test papers in the Appendix.

These tests align closely with the Gambian national curriculum for grades 1-3. Versions of them have also been used as part of the government's efforts to assess its own teachers since 2007. This ensures that our measures of learning hew closely to the education goals of the Gambian national education system.

We generate four measures of learning using these tests. First, we generate a composite score of overall child performance at endline, calculated as the proportion of total questions answered correctly on each of the two tests.[11] We estimate both the difference in (raw) composite scores between groups, as well as the transformation of this difference into standard deviation units using Cohen's d. We refer to this as our "SD" measure.

We also study children's acquisition of specific skills, using performance on the individual sub-

[11]Each test is given equal weight in generating this measure, and within each test, performance on each subtask is given equal weight.

tasks within each test. First, we use binary variables capturing whether the child meets established thresholds for literacy and numeracy (Dubeck and Gove, 2015; Fazzio et al., 2021). A child is assessed to be literate if they can read "with good fluency" (45 words per minute; subtask 5a) and correctly answer at least 80% of reading comprehension questions (subtask 5b). A child is assessed to be numerate if they can successfully identify missing numbers in a sequence (e.g., 2, 4, _, 8) in at least 70% of the questions on the test (subtask 3), and correctly answer at least 80% of word problems (subtask 6). Finally, we study differences in child performance on each of the individual subtasks, as measured by the proportion of questions in that subtask answered correctly. This final set of measures allows us to show detailed learning trajectories across a spectrum of skills, from the very earliest stages of letter and number recognition to more advanced skills on the path to literacy and numeracy.

In consultation with the Gambian Ministry of Basic and Secondary Education and other experts in the area, at the end of pre-trial fieldwork we decided not to conduct baseline tests of learning. Our fieldwork suggested that, because our focus was on children who had not yet been to school at the time of baseline enumeration (and prior to randomization), baseline tests would have generated only a trivially small number of non-zero scores, and therefore the cost – both financial and in terms of the time and energy of participants – greatly exceeded the likely benefit of these tests. We assume every child starts from a zero baseline learning level in terms of the skills we measure at endline; the very low levels of these skills that we measure in the status quo group, after the vast majority of students have completed three years of primary schooling, support this assumption.

2.5 Measuring demand via aspirations

We use two measures to proxy for the following latent traits of each family: their desire to obtain a better adult quality of life for their child than experienced by previous generations, achieved via either schooling or employment, respectively. The goal of these questions was to capture variation in these traits, both within and across families, and to see how this variation related to subsequent investment behavior and differences in endline learning.

These data were collected at baseline, prior to randomization and before study children entered school for the first time. The questions asked the child's main caregiver about their aspirations for the child's future schooling and employment. They were piloted prior to use, and are similar to those asked in other studies of aspirations in Ethiopia, India, and Somalia (Bernard et al., 2014; Attanasio et al., 2020; Kipchumba et al., 2021).

Following La Ferrara (2019), we measure two types of aspiration.[12] The first is the caregiver's aspirations for their child's highest level of educational attainment. To capture these "educational aspirations," we asked the child's main caregiver: "ideally, what is the highest level of education you would like [child name] to attain?"[13] The second is the caregiver's aspirations for their child's career in adulthood. To capture these "career aspirations," we asked the caregiver: "when [child name] is 20 years old, what job do you hope [she/he] will be doing?" We transform these into binary variables. For educational aspirations, we generate an indicator variable for whether the caregiver would like the child to attend university. For career aspirations, we generate an indicator for whether the caregiver hopes the child will work in an urban area[14], capturing the fact that most jobs in urban areas require literacy and numeracy skills, and on average pay substantially more than jobs in the countryside.

In Table 3, we present average values and conditional means of aspirations levels at baseline. We show these conditional means by treatment status and by a series of variables related to relative economic prosperity, household features, and caregiver education, all of which are predetermined relative to measurement of aspirations. Around 60 percent of caregivers would like their child to go to university, which we call "high" educational aspirations. This is slightly lower than lev-

[12]The aspirations we measure differ importantly from expectations. In our pilot, we attempted to measure both, and worked to choose language that differentiated between aspirations and expectations. In this work, however, we determined that we could not ask respondents about both expectations and aspirations without unacceptably large priming effects.

[13]Lybbert and Wydick's 2018 study of aspirations differentiates between "aspirational hope" and "wishful hope," arguing that the latter is characterized by a lack of a viable pathway to achieve the desired outcome. Among our study participants, as in the Ethiopian, Indian, and Somalian contexts referenced above, few individuals are likely to go to university. Nonetheless, many caregivers hope that their children will do so, and we follow this body of prior research in referring to responses to the education question as capturing educational aspirations. As we show later, they also serve as predictors of investment in education; in other words, whether the reader sees them as aspirational hope or wishful hope, they perform well as measures of latent family demand.

[14]This includes jobs such as doctor, nurse, judge, law clerk, or politician, but not imam, farmer, or farm laborer.

els recorded (roughly) contemporaneously in rural Ethiopia (Bernard et al., 2014) and Somalia (Kipchumba et al., 2021), and far lower than measured a few years earlier in India (Attanasio et al., 2020). Around 65 percent of caregivers aspire that their child will work an urban area, what we call high career aspirations. The correlation between educational and career aspirations is 0.181, indicating substantial independent variation between the two. We see no difference in baseline aspirations between the caregivers of children in the intervention and status quo groups.

Two stylized facts emerge from Table 3. First, baseline aspirations are higher for some baseline characteristics that might predict educational investment and learning levels (caregiver education and literacy), but these associations are less strong for other traits (i.e., wealth[15]). Second, there is substantial variation in aspirations independent of these variables. For example, even among caregivers with no formal schooling, and separately for those who cannot read, nearly 60 percent also express high educational and career aspirations for their children. This variation is also evident in the correlation between the variables and aspirations, which is positive but uniformly small in magnitude, ranging from 0.012 to 0.127. In our analysis of the relationship between baseline aspirations and subsequent educational investment and learning gains, we control for these variables, isolating the relationship between our dependent variables and the part of our aspirations measures which are orthogonal to these other independent variables. A further robustness test in response to the concern of overlap between aspirations and other family traits shows that our findings are also robust to dropping the eight percent of observations in which the caregiver is literate at baseline, the variable with which aspirations are most highly correlated.[16]

We also conduct and report a random forest analysis of the importance of all of our independent variables in predicting performance on endline learning levels in the status quo group. Using the *rforest* Stata command and conducting our analysis separately for each type of aspiration, we find that aspirations are the second most-important predictor of endline learning, after only caregiver literacy. The third, fourth, and fifth most important predictors are books found in home, household

[15]This reflects the fact that, in rural parts of The Gambia, higher levels of wealth are not necessarily predictive of greater education, particularly given the importance of farming and animal husbandry.
[16]Not included here for length; available on request.

Table 3: Levels of aspirations at baseline, overall and conditional means

	(1) Aspires that child will go to university	(2) Correlation of variable with educational aspirations	(3) Aspires that child will go to university	(4) Correlation of variable with career aspirations
Overall	0.61		0.65	
Randomization group				
Intervention	0.61	0.006	0.65	0.003
Status quo	0.61		0.65	
P-value of difference	(0.72)		(0.87)	
Child gender				
Male	0.63	0.030	0.64	-0.033
Female	0.60		0.67	
P-value of difference	(0.07)		(0.04)	
Caregiver education				
Caregiver has been to school	0.71	0.114	0.74	0.104
Caregiver has never been to school	0.58		0.63	
P-value of difference	(0.00)		(0.00)	
Caregiver literacy				
Can read simple sentence	0.82	0.127	0.82	0.106
Cannot read simple sentence	0.59		0.64	
P-value of difference	(0.00)		(0.00)	
Materials of home				
Home made of synthetic materials	0.62	0.012	0.68	0.034
Home made of natural materials	0.61		0.64	
P-value of difference	(0.47)		(0.04)	
Books in house				
Books found in house	0.63	0.046	0.67	0.041
No books found in house	0.58		0.63	
P-value of difference	(0.00)		(0.01)	

Table 3 notes: this table shows the unconditional mean for each of the two aspirations we study and their conditional means by each of the binary baseline characteristics labeled in italics in the left-most column. For conditional means, we conduct a t-test of the null that the aspiration in question is equal for those with each value of the baseline characteristic, and present the p-value in parentheses below. We also present the pairwise correlation between the variable listed and each aspiration. Caregiver literacy is an indicator for whether the caregiver can read a simple sentence – in the spirit of the ASER literacy test (Pratham, 2012) – at the time of a baseline survey. The household wealth variable is described in the text. Books in house is indicator for whether there were any books found in the child's home during the baseline survey.

wealth, and caregiver education.

We acknowledge that there is a vast range of of socio-economic characteristics which could be correlated with both educational aspirations and test scores, and that we only analyze a limited set of these. Our fieldwork suggested that the variables we use were by far the most important measures of latent SES, however, and this informed our choice of their inclusion. Furthermore, the extremely low levels of wealth among people living in theses areas suggest that there is less variation in SES in our study area than in even moderately wealthier settings where aspirations have previously been studied (Lybbert and Wydick, 2018; Ross, 2019).

We argue that these aspirations measures capture (part of) latent family demand for investment in their children's future. Two features of our data support this argument. First, as we show in Section 4, these measures are significant predictors of subsequent investment in the child's education. Second, as we show in Section 5, they appear to be family-specific rather than child-specific. Among the families with multiple children in our study, between 70 percent (career) and 90 percent (education) report the same aspiration for both children. This suggests we are likely capturing family demand, rather than traits of the child such as unobserved ability. In Section 5.3, we discuss these issues in greater depth.

3 Research design

Our empirical analysis aims to estimate two relationships.[17] The first is the relationship between educational demand measured at baseline and subsequent educational investment and early learning outcomes. We estimate this relationship using the following equation:

$$y_{ic} = \alpha_0 + \alpha_1 A_{t=0,ic} + \alpha_2 X_{t=0,ic} + \eta_r + \varepsilon_{ic} \tag{1}$$

In this equation, y_{ic} is the outcome variable for child i in cluster c; α_0 is a constant; $A_{t=0,ic}$ is aspirations of the caregiver for child i at baseline (i.e., when $t = 0$); $X_{t=0,ic}$ is a vector of predetermined

[17]While the analysis for the RCT reported in Eble et al. (2021) was pre-specified and pre-registered (Boone et al., 2015), this paper reports exploratory analysis for which we chose not to pre-register an analysis plan (Olken, 2015; Lin and Green, 2016).

variables for child i, measured at baseline, which include all the variables shown in Table 3; and η_r is a region-specific fixed effect. We cluster our standard errors, ε_{ic}, at the level of the cluster of contiguous villages, the level at which randomization was conducted.

Our main parameter of interest is α_1, which captures the mapping from baseline demand, as measured by aspirations, to subsequent outcomes. This estimate is conditional on the region of the child's village and the baseline characteristics contained in $X_{t=0,ic}$ (and listed in Table 3), such as gender, wealth, and caregiver education.[18] Our approach to estimating this parameter is not causal; rather, it captures the observational relationship between our measure of latent demand and two objects of interest to many economists: investment and skill acquisition.

To estimate α_1, we use only data from the status quo group. This is because the intervention group's subsequent educational investment and endline learning levels are affected by receipt of the intervention, confounding our ability to measure the status quo mapping from baseline aspirations to subsequent outcomes among children in this group.[19] We address remaining sources of omitted variable bias in this estimate in Section 4.4.

Second, we generate causal estimates of how a dramatic increase in the quality of educational supply changes the mapping from baseline demand to endline learning. To do so, we study children in both the status quo and intervention groups, using random assignment of the bundled para teacher intervention as a source of identifying variation. This estimation also uses ordinary least squares, regressing the outcome variable on a constant, baseline aspirations, the randomly assigned treatment status of the village in which the child was enumerated, T_c, and their interaction, using the same set of controls and error clustering strategy as in Equation 1:

$$y_{ic} = \beta_0 + \beta_1 A_{t=0,ic} + \beta_2 T_c + \beta_3 T_c * A_{t=0,ic} + \beta_4 X_{t=0,ic} + \eta_r + \varepsilon_{ic} \tag{2}$$

Our main parameter of interest from this equation is β_3. The sign and significance of β_3 indi-

[18]For clarity of exposition, we use only the variables listed in Table 3 in our main specification. All results are robust to including the remaining controls - such as language and ethnicity fixed effects - used in other analyses of these data (Eble et al., 2021).

[19]For completeness, in Appendix Table A.1 we show these relationships for both the status quo and intervention groups, estimated using Equation 2.

cate whether the change in the quality of educational supply induced by the intervention changes the relationship between baseline demand and endline learning. A positive and significant estimate of β_3 would suggest that family demand and educational supply are complementary, while a negative and significant estimate would suggest substitutability between the two, including possible substitution behavior on the part of the family. β_1 in this equation is analog to α_1 from the Equation 1; β_2 captures the overall effect of the intervention.

We also estimate a parameter which we call the "interaction mean." This captures the mean level of the outcome variable for high-aspirations children, conditional on being enumerated at baseline in a village that was later randomly assigned to receive the intervention. We calculate this by adding β_1 and β_3. In our results we present a p-value of a test of the null that the interaction mean is equal to zero. The magnitude and statistical significance of the interaction mean are estimates of whether, in intervention villages, children of high-aspirations caregivers demonstrate a higher endline level of the skill in question than do children of other families in the same village.

4 Demand, investment, and learning in the status quo

In this section, we characterize the mapping from baseline demand, as measured by aspirations, onto subsequent educational investments and endline learning levels in the rural Gambian status quo. We estimate Equation 1 using the measures of investment and learning described in Section 2. We then bound our results by describing the likely sign of any potential influence from unobserved factors on our estimates.

4.1 Demand and educational investment in the status quo

We first characterize the mapping from baseline aspirations levels to subsequent educational investment, presenting our estimates in Table 4. The outcome variables, given in the column headings, are family educational expenditure on the child in year three of the study, child time use in year three of the study, and the child's enrollment in school in each of the three study years.

Baseline aspirations have a positive and statistically significant mapping onto subsequent educational investments. Caregivers who hold high educational or career aspirations for the child spend between 10 and 15 percent more money per year on costs related to the child's education than other caregivers.[20] Children of these caregivers also spend a greater proportion of their time on a typical weekday on school-related tasks. This difference in time use is statistically significant for baseline educational aspirations, but not for baseline career aspirations.

Children of high-aspirations caregivers are also more likely to be enrolled in school in the first two years of the study. This pattern disappears in year three of the study, at which point almost all children are enrolled in school. Nonetheless, this early difference is important: the greater likelihood of delayed enrollment among children of low-aspirations caregivers suggests lower expected educational attainment for these children (Nonoyama-Tarumi et al., 2010).

As a check for plausibility, we compare the sign and magnitude of our estimate of α_1 to similar relationships in this setting, as well as to estimates from another, similar setting. Our estimates of the mappings from the control variables to educational investment – also shown in Table 4 – have a similar order of magnitude as do those for baseline aspirations, and the signs of these estimated relationships are as expected: there is, for example, a statistically significant positive relationship between wealth and educational expenditure. Aspirations also survive the "horse race" between this set of control variables in predicting these outcomes. Second, we note that estimates of the impact of an intervention-driven aspirations gain on educational investment in rural Ethiopia (Bernard et al., 2014) are similar in sign and magnitude to our estimates of the mapping from aspirations to investment in rural Gambia.

4.2 Demand and learning in the status quo

We next estimate how baseline aspirations map onto endline learning levels in the status quo group. We present our first set of results in Table 5. In column 1 we show this relationship for raw test scores. At endline, children whose caregivers expressed high baseline educational aspirations for the child perform 3.3 points better than other children, from a comparison group mean of 15

[20]Expenditures are reported in Gambian Dalasis. In mid-2018 when these data were collected, the exchange rate between Dalasis to US Dollars was 46.81 to one.

Table 4: Baseline aspirations and educational investment in the status quo

	(1) Educational expenditure	(2) School-related time use	(3) Enrolled in school, year 1	(4) Enrolled in school, year 2	(5) Enrolled in school, year 3
Panel A: Educational aspirations					
Aspiration: child will go to college (α_1)	76.70**	0.019***	0.031	0.055**	0.006
	(27.88)	(0.007)	(0.027)	(0.025)	(0.008)
Wealth index high	122.44***	0.003	-0.008	-0.027	-0.001
	(41.19)	(0.007)	(0.024)	(0.018)	(0.012)
Caregiver can read simple sentence	80.58	0.025**	0.063*	0.049*	0.016*
	(73.77)	(0.012)	(0.034)	(0.026)	(0.008)
Books found in house	62.97**	0.008	0.051**	0.040***	0.005
	(30.64)	(0.007)	(0.021)	(0.013)	(0.006)
Child is female	-13.58	0.006	0.018	-0.000	0.010
	(23.67)	(0.008)	(0.015)	(0.022)	(0.008)
Comparison group mean	611.36	0.545	0.825	0.802	0.971
Number of observations	1,923	1,970	2,002	1,970	1,970
Panel B: Career aspirations					
Aspiration: child will work in urban area (α_1)	69.25**	0.005	0.034	0.055***	0.000
	(27.44)	(0.006)	(0.023)	(0.020)	(0.005)
Wealth index high	119.58***	0.003	-0.009	-0.029	-0.001
	(40.44)	(0.007)	(0.024)	(0.018)	(0.012)
Caregiver can read simple sentence	83.81	0.027**	0.064*	0.051*	0.017**
	(71.22)	(0.012)	(0.035)	(0.028)	(0.008)
Books found in house	67.10**	0.009	0.053**	0.043***	0.005
	(30.20)	(0.007)	(0.020)	(0.013)	(0.006)
Child is female	-19.61	0.005	0.016	-0.005	0.010
	(23.58)	(0.008)	(0.015)	(0.023)	(0.008)
Comparison group mean	617.54	0.553	0.820	0.799	0.973
Number of observations	1,923	1,970	2,002	1,970	1,970

Table 4 notes: this table reports the results of estimating Equation 1 using the outcome variable given in the column heading and the baseline aspirations (educational or career) indicated in the panel heading. Dependent variables labeled in the column headings are defined in the text. These analyses include only children in the status quo group. We report clustered standard errors in parentheses below each estimated coefficient. Observations vary by column because outcome variables were collected at different times and some children were missed in some periods. Results are robust to including only the smallest estimation sample. The full set of controls is as indicated in Section 3. $^{*}p < 0.10$, $^{**}p < 0.05$, $^{***}p < 0.01$. For completeness, in Appendix Table A.1 we show these relationships for both the status quo and intervention groups, estimated using Equation 2.

Table 5: Baseline aspirations and endline learning in the status quo

	(1) Endline test score	(2) Child is literate	(3) Child is numerate	(4) Words read per minute
Panel A: Educational aspirations				
Aspiration: child will go to college (α_1)	3.390***	-0.001	-0.002	1.147**
	(0.942)	(0.002)	(0.005)	(0.507)
Wealth index high	1.821*	0.001	-0.002	1.252*
	(1.027)	(0.002)	(0.004)	(0.645)
Caregiver can read	5.937***	-0.001	-0.005	1.380*
	(1.420)	(0.001)	(0.004)	(0.791)
Books found in house	2.678***	0.001	-0.001	0.449
	(0.705)	(0.001)	(0.002)	(0.347)
Child is female	1.746**	-0.002	0.000	0.256
	(0.866)	(0.001)	(0.003)	(0.354)
Comparison group mean	14.964	0.001	0.006	1.991
Number of observations	2,039	2,039	2,038	2,033
Panel B: Career aspirations				
Aspiration: child will work in urban area (α_1)	3.603***	0.002	0.001	1.268***
	(0.633)	(0.001)	(0.003)	(0.339)
Wealth index high	1.696	0.001	-0.002	1.207*
	(1.043)	(0.002)	(0.004)	(0.658)
Caregiver can read	6.018***	-0.001	-0.005	1.401*
	(1.420)	(0.001)	(0.005)	(0.768)
Books found in house	2.887***	0.001	-0.001	0.521
	(0.690)	(0.001)	(0.002)	(0.339)
Child is female	1.466	-0.002	0.000	0.158
	(0.893)	(0.001)	(0.003)	(0.354)
Comparison group mean	14.604	0.000	0.004	1.806
Number of observations	2,039	2,039	2,038	2,033

Table 5 notes: this table reports the results of estimating Equation 1 using the outcome variable given in the column heading and the baseline aspirations (educational or career) indicated in the panel heading. Dependent variables labeled in the column headings are defined in the text. These analyses include only children in the status quo group. We report clustered standard errors in parentheses below each estimated coefficient. The scale of the endline test score is 0-100. Literacy and numeracy are indicator variables. The full set of controls is as indicated in Section 3. $^{*}p < 0.10$, $^{**}p < 0.05$, $^{***}p < 0.01$.

points.[21] For children of caregivers with high career aspirations, this difference is 3.8 points. Both differences are highly statistically significant.

We plot the distribution of these scores, by aspiration group, in Figure 1. This shows that, for both types of aspiration, the high-aspirations group's test score distribution first-order stochastically dominates that for children of other families. Kolmogorov-Smirnov tests of equality of distributions reject equality with $p < 0.001$ in both cases.

Using the common practice of transforming raw score differences into standard deviation units, the mapping from baseline caregiver aspirations to endline learning appears very large. For educational aspirations, the raw difference translates into a difference of 0.28 SD, and for career aspirations, it is 0.30 SD.[22] A series of recent meta-analyses summarize estimates from hundreds of evaluations of educational interventions in such settings (c.f. McEwan, 2014; Glewwe and Muralidharan, 2016; Evans and Yuan, 2022). In the context of these studies, an intervention with an effect size of 0.28–0.30 SD would lie between the 75th and 90th percentile of all known estimates.

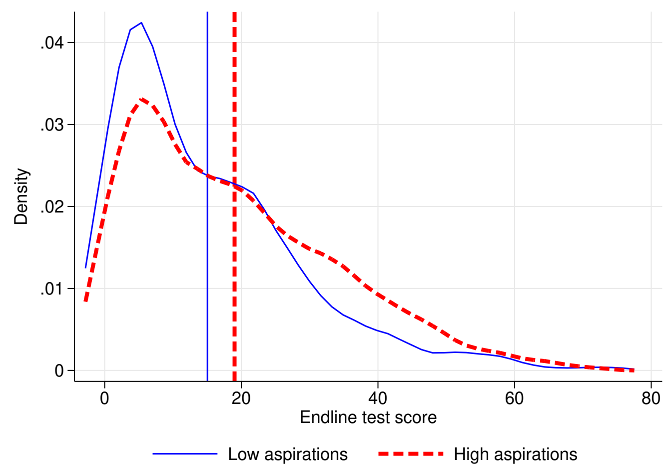
Measuring learning with skill acquisition, rather than the SD, paints a far different picture. We show estimates of the same demand–learning relationship in columns 2 and 3 of Table 5, using our binary measures of basic literacy and numeracy as the outcome variable, respectively. We find that children of high aspirations caregivers are no more likely to reach either literacy or numeracy than other children, estimating precise zeroes in both cases: the confidence intervals around our estimates reject anything larger than a one percentage point difference.[23] In column

[21] In other words, the average child not in the high-aspirations group correctly answers 15 percent of questions correctly; this is shown in the "comparison group mean" row.

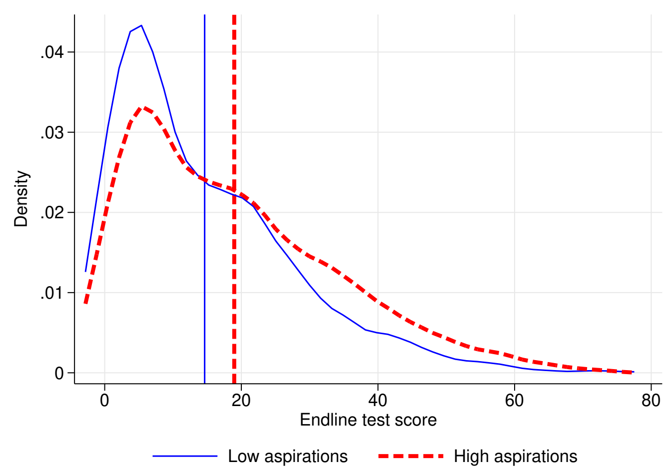
[22] Estimated using *Cohen's d*.

[23] Hundreds of studies, as well as several meta-analyses, use effect sizes stated in SD terms for comparison of the strength and magnitude of the relationship between various educational inputs and learning outcomes (Kremer and Holla, 2009; McEwan, 2014; Ganimian and Murnane, 2016; Glewwe and Muralidharan, 2016; Evans and Yuan, 2022). Our findings here – in particular the comparison between our estimates when using the SD measure as our dependent variable, as opposed to estimates when using skill-based measures of learning – show that in cases where learning levels are very low, using the test score SD metric to compare across contexts can lead to overly optimistic conclusions about the relative importance of different inputs. This is primarily because low levels of baseline variation (i.e., due to the compression of the distribution of scores near zero) make small absolute gains appear as large relative gains. This underscores the conclusions of prior work outlining the psychometric issues with the comparability of different tests, and particularly the problems with using the SD measure that these studies point out (Hill et al., 2008; Kraft, 2020; Furr, 2021; Evans and Yuan, 2022). It also suggests that, in such settings and for cross-context comparison, measures of absolute skill acquisition should be preferred.

Figure 1: Distributions of endline test scores in the status quo, by baseline aspirations



Panel A: Educational aspirations



Panel B: Career aspirations

Figure 1 notes: this figure shows kernel density plots of endline test scores for children whose caregivers did (red dashed line) and did not (solid blue line) express the aspiration listed in the panel title at baseline. In these plots, we focus on children in the status quo group (that is, in villages assigned to not receive the intervention) and for whom we have a test score, comprising 1,971 observations. The vertical lines show the mean test score of the group whose distribution is plotted using the same width, color, and pattern of line. Kolmogorov-Smirnov tests reject the equality of the two distributions with $p \leq 0.001$ in each panel.

4 we show results for a related skill, correct words read per minute. Here we see that children of high-aspirations caregivers can read roughly one additional word per minute, from a comparison group mean of less than two total words read of the 50 total words on the test (put in SD terms, this would be a 0.14 SD difference). For reference, a common benchmark for reading proficiency is reading between 45 and 60 words per minute (Dubeck and Gove, 2015).

Regardless of aspirations, essentially zero children in the status quo group are anywhere near literacy and numeracy at endline. A visual rule of thumb is that children approach literacy and numeracy when they can correctly answer between 60 to 65 percent of the questions on these two tests. Applying this to the distributions in Figure 1, and seeing these alongside the other results from this section, it is clear that demand alone is insufficient to reach meaningfully higher learning levels in this, and perhaps similar settings.

4.3 Estimates for the acquisition of individual skills

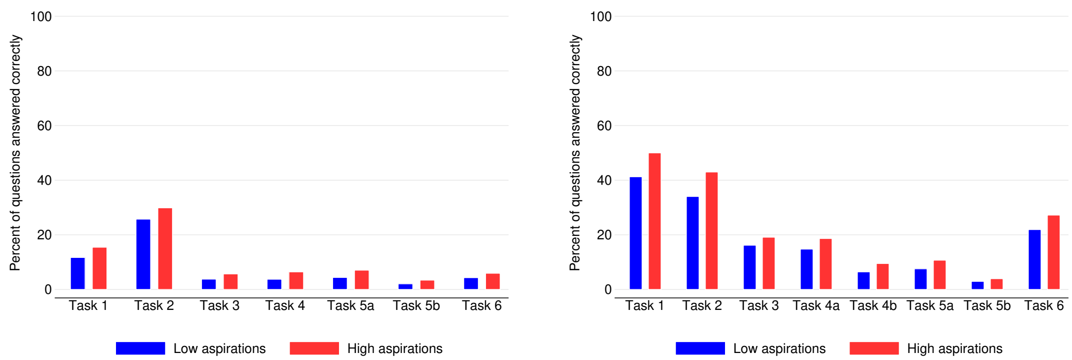
We next estimate how baseline aspirations map onto the acquisition of lower-level reading and math skills that precede literacy and numeracy. As described in Table 2, each test comprises a series of subtasks that evaluate different sets of skills, such as number and letter recognition, familiar word recognition, and single-digit addition. In Figure 2 we present the average proportion of questions in each subtask that children in status quo villages answered correctly. We show this separately for each type of aspiration. In Tables A.2 and A.3, we show regression results for this comparison, estimating Equation 1 using the relevant subtask score as the dependent variable.

These analyses reveal two main facts about children's early grade learning in the rural Gambian status quo. First, after two to three years of schooling, endline skill levels are extremely low regardless of baseline aspirations. For most math and reading skills – such as single-digit subtraction or the ability to read simple, familiar words such as "and" and "but" – children in both groups correctly answer fewer than 10 percent of questions.

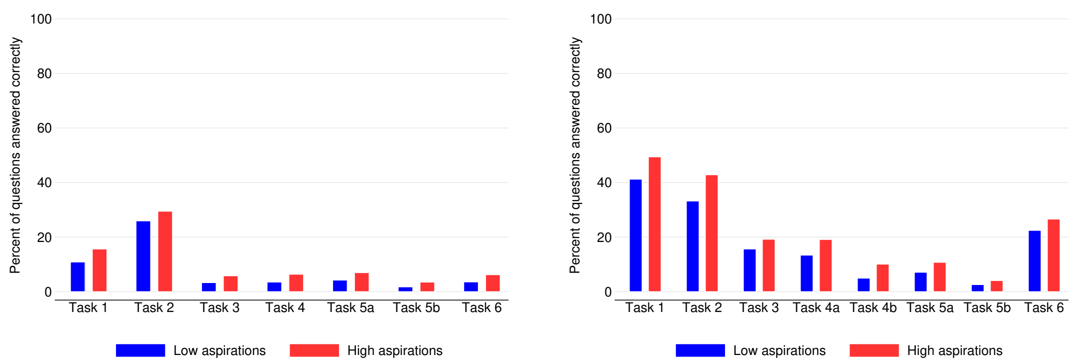
Second, even though the SD measure shows large relative differences in performance between children of high-aspirations caregivers and other children, the absolute differences in skill levels

Figure 2: Endline skill levels in the status quo group, by baseline aspirations

Reading skills *Math skills*



Panel A: Educational aspirations



Panel B: Career aspirations

Figure 2 notes: this figure shows endline performance, by baseline aspirations level, on each of the individual subtasks of the EGRA and EGMA tests, respectively. Panel titles indicate the aspiration being studied. In these plots, we focus on children in the status quo group (that is, in villages assigned to not receive the intervention) and for whom we have a test score, comprising 1,971 observations. The subtasks listed on the x-axis are described in Table 2 and the full test papers are given in Appendix A.

between the groups are extremely small. This is illustrated in Figure 2, Tables A.2 and A.3, and the results for words per minute (wpm) in Table 5. Endline wpm is 50% higher for children of caregivers with high educational aspirations than for others. Because their endline wpm rate is fewer than two, however, these children are observationally equivalent to other status quo children in terms of their very low likelihood of ever reaching foundational literacy or numeracy.

This is particularly troubling for the age group of children we study. At the end of this study, these children are between nine and 12 years old, and three quarters of them are in the second or third grade. As they progress to higher grades, the school curriculum will advance from teaching the encoding and decoding skills that comprise literacy and numeracy to more abstract skills which themselves rely upon mastery of literacy and numeracy. Given how far the students in our study population are from mastering these skills, they are extremely likely to be left behind as school progresses. As a result, they are highly unlikely to ever attain the skills comprising either basic literacy or numeracy in their schooling careers (Cunha and Heckman, 2007; Pritchett, 2013; Pritchett and Beatty, 2015; Muralidharan et al., 2019).

This mirrors findings from psychology in the field of child development. The absence of critical inputs during this period can function as what the developmental literature refers to as "a violation of the expectable environment," or the "absence of an expected experience" (Nelson III and Gabard-Durnam, 2020, p. 134). These harms have long-lasting knock-on effects, rendering it difficult for the child to ever acquire the skill in question – in this case, literacy and numeracy. Because these two skills are prerequisites for the attainment of many other, higher-level skills, the main consequence of this breakdown in the learning process is a very low expected ceiling for their subsequent learning trajectory.

4.4 Bounding our estimates

We argue that our estimates are a likely upper bound on the true relationship between family demand, educational investment, and child learning for these children. Aspirations for education and employment are often positively correlated with other hard-to-measure or unobservable traits

– such as caregiver wealth, education, or other tastes and preferences – that are also positively correlated with child educational investment and outcomes (Bernard et al., 2014; Ross, 2019). As a result, any confounding from such sources would cause our estimates to be exaggerated, relative to the true relationship (Wooldridge, 2016). Therefore, unless there exists some other influential but unobserved trait which is negatively correlated with these specific aspirations and positively correlated with educational investment and learning outcomes (or vice versa), our estimates are likely to be larger in magnitude than the true relationship between demand and learning.

Extrapolating beyond our sample to the population of all children in these areas, we argue that our estimates are also an upper bound on the relationship between family demand and learning for this broader group. As described in Section 2.3, presence in our sample is conditional on the caregiver intending to enroll the child in school in the coming year.[24] For children of eligible age, but whose caregivers did not intend to send them to school in the coming year, we argue that our estimates of α_1 in Table 5 are also an upper bound on the learning differentials between children of high-aspirations caregivers and other children. This is because the children excluded by this inclusion criterion are likely to have either a similar or worse learning trajectory than study participants, given the negative consequences of delayed school enrollment for learning and schooling (cf. Glewwe and Jacoby 1995; Bommier and Lambert 2000).

5 How demand and supply interact to produce learning

In this section, we estimate how a dramatic increase in the quality of educational supply changes the mapping from demand to learning estimated in Section 4. We then provide evidence on the substitutability and complementarity of demand and supply in generating learning at different levels of skill. Our analysis in this section reports estimates of Equation 2 using data from the entire sample, i.e., both the status quo and intervention groups. We focus on β_3, which captures the interaction between baseline aspirations and the large change in the quality of educational supply

[24] In our sample, this eligibility criterion excluded roughly 13 percent of children at baseline who would otherwise be eligible according to our two remaining eligibility criteria: one, the child's age; and two, their not having previously attended school at grade 1 or higher.

caused by the randomly-assigned intervention.

5.1 Demand, supply, and learning

We show results in Table 6 using the four summary learning outcomes studied in Section 4: standardized test scores, literacy, numeracy, and correct words read per minute. In Panel A, we show these results for educational aspirations; in Panel B, we show them for career aspirations.

In the presence of high-quality educational supply, we find a positive, large, and statistically significant mapping from baseline educational aspirations to endline learning. This stands in stark contrast to the zero mapping from either aspiration to learning shown in Section 4. While our estimates in Column 1 show no evidence of an interaction effect in our test score measure, our estimates in Columns 2, 3, and 4 show that, conditional on receiving the dramatic improvement in the quality of educational supply provided by the intervention, baseline educational aspirations map onto significantly greater acquisition of high-level reading and math skills. For literacy, children of caregivers with high educational aspirations are six percentage points more likely to achieve literacy (from a comparison group mean of 23 percent) and they are four percentage points more likely to achieve numeracy (from a comparison group mean of 17 percent). These comprise a roughly 25 percent increase in the child's likelihood of achieving each of these levels of reading and math ability at endline. These children can also read more than three extra words per minute (from a comparison group mean of 35), or a roughly 10 percent increase. While this is a threefold increase in the magnitude of mapping from baseline aspirations to wpm at endline (comparing the magnitudes of β_3 to α_1), more important still is *where* in the support of possible wpm that this gain happens. Going from 35 to 38.5 words per minute suggests faster comprehension of written material and the potential for more rapid learning. In the status quo group, on the other hand, the difference between being able to read 1.8 and 3 words per minute is not likely to accelerate learning in a similar manner. This is particularly true if these students advance to higher grades, where teaching to these students' appropriate level is even less likely (Banerjee et al., 2017a).

We interpret these results – no evidence of an effect on raw test scores, but positive, significant, and large estimated effects for literacy, numeracy, and words per minute – as evidence that the

Table 6: How the mapping from baseline aspirations to endline learning changes with a large increase in the quality of educational supply

	(1) Endline test score	(2) Child is literate	(3) Child is numerate	(4) Words read per minute
Panel A: Educational aspirations				
Aspirations x intervention (β_3)	0.39	0.06***	0.04*	3.15**
	(1.58)	(0.02)	(0.02)	(1.59)
Intervention (β_2)	45.52***	0.23***	0.17***	35.22***
	(1.74)	(0.02)	(0.02)	(1.77)
Aspirations (β_1)	3.65***	-0.00	-0.00	1.17***
	(0.92)	(0.00)	(0.01)	(0.49)
Interaction mean ($\beta_1 + \beta_3$)	4.04	0.06	0.04	4.32
P-value [$\beta_1 + \beta_3 = 0$]	[0.002]	[0.019]	[0.081]	[0.005]
Comparison group mean	14.96	0.00	0.01	1.99
Number of observations	3,814	3,814	3,813	3,805
Panel B: Career aspirations				
Aspirations x intervention (β_3)	-2.44*	0.03	0.01	0.87
	(1.32)	(0.02)	(0.02)	(1.27)
Intervention (β_2)	47.33***	0.25***	0.18***	36.56***
	(1.68)	(0.03)	(0.02)	(1.75)
Aspirations (β_1)	3.86***	0.00	0.00	1.20***
	(0.64)	(0.00)	(0.00)	(0.35)
Interaction mean ($\beta_1 + \beta_3$)	1.42	0.03	0.01	2.07
P-value [$\beta_1 + \beta_3 = 0$]	[0.216]	[0.162]	[0.595]	[0.093]
Comparison group mean	14.60	0.00	0.00	1.81
Number of observations	3,814	3,814	3,813	3,805

Table 6 notes: this table reports our estimates of the parameters in Equation 2 for the outcomes listed in the column headings. The panel titles indicate which baseline aspirations were used to generate the estimates shown. Coefficient estimates are reported according to the row title. We report clustered standard errors in parentheses below each estimated coefficient. Each panel-by-column "cell" corresponds to a separate regression. Comparison group means are calculated for those in the status quo group whose caregiver did not express the aspiration given in the column title at baseline. $^{*}p < 0.10$, $^{**}p < 0.05$, $^{***}p < 0.01$.

increase in the quality of educational supply shifted the role of family demand. In the status quo, family demand is associated with sizable but developmentally inconsequential improvements on extremely low learning levels. High quality educational supply moves *all* children further along in the process of skill acquisition, at which point the demand–learning relationship we measure in Section 4 does indeed make a developmentally meaningful difference for skill acquisition. We explore this further in Section 5.2.

For children in intervention villages whose caregivers express high career aspirations at baseline, we observe a different pattern of results. Our estimate of β_3, the relationship between baseline career aspirations and endline test scores in the face of high-quality educational supply, is negative and statistically significant. This suggests possible substitution away from child learning in the face of the increase in the quality of educational supply. For the three other learning outcomes – literacy, numeracy, and words read per minute – our estimates of β_3 are much smaller in magnitude than for educational aspirations, and none are statistically distinguishable from zero.

We can also show these results graphically. In Figure A.2, we plot the distribution of test scores among the four relevant groups – children with high-aspirations caregivers and other children; and those born in villages who were randomized to (not) receive the intervention. As in Figure 1, we show separate panels for educational and career aspirations. Here, similarly, Kolmogorov-Smirnov tests strongly reject equality of the test score distributions for children of high aspirations caregivers and other children, respectively, for three of the four combinations of aspirations type (educational or career) and intervention group (status quo or intervention) with ($p < 0.001$). The exception is for career aspirations in intervention villages, where we cannot reject equality ($p > .10$).

We next report sensitivity analyses for these results. The literacy and numeracy variables, while coded based on accepted levels of skill mastery for these two tests, are binary. In Figure A.3, we show how sensitive our results are to alternative specifications of literacy and numeracy based on other, arbitrary thresholds for performance on the component skills comprising each measure. This figure reports a heat map of estimates of β_3 from Equation 2, using 10,000 alternative, arbitrary "pseudo-" measures of literacy and numeracy, consisting of each location on the 100-by-100 unit grid of all possible integer thresholds for the percent of questions answered correctly on each of the two subtasks comprising each literacy and numeracy. For clarity of exposition, we display all estimates with $\beta_3 \leq 0$ or with $p \geq 0.10$ as white space.

This analysis shows that our main results are robust across a range of potential thresholds in the neighborhood of the accepted standards for defining literacy and numeracy. Furthermore, in many cases our estimates would be larger in magnitude were we to choose another, slightly more lenient threshold. In addition, a key pattern we see in Table 6 appears here as well – strong evidence of a positive interaction between baseline educational aspirations and educational supply in generating learning at endline, and far weaker evidence of an interaction between baseline career aspirations and educational supply.

5.2 How demand and supply combine in the production of skills

In this section, we estimate how demand and supply combine to produce the skills that lead up to, and later form foundational literacy and numeracy. The tests we use are comprised of groups of questions that capture the mastery of different skills, as described in Table 2. These span from some of the most basic testable skills, such as letter and number recognition, to far harder ones, such as reading comprehension and two-digit subtraction with borrowing.

We report estimates of Equation 2 using child performance on these different subtasks in reading and math as outcome variables. The sign and significance of our estimates for β_3 capture this interaction between demand and supply in the acquisition of different levels of skill. Recall that negative estimates of β_3 indicate substitutability between them, and positive estimates indicate complementarity.

We present our results in graphical form for educational aspirations in Figure 3, and for career aspirations in Figure 4. These plot our point estimates and a 95 percent confidence interval for β_3, estimated separately for each skill. We plot these estimates in ascending order of skill difficulty within each test (reading or math), with the most difficult skills at the right-most place in the figure.

We find evidence of substitutability and complementarity between demand and supply that vary systematically as the difficulty of the skill being acquired increases. These patterns also vary by aspiration type. For educational aspirations (Figure 3), we estimate a positive gradient between β_3 and skill difficulty. For performance on lower-level subtasks (1 and 2 in both reading and math) we generate negative – and for math, statistically significant – estimates, suggesting substitutability between supply and demand. As skill difficulty increases, the estimates become positive. For the most difficult subtasks – reading and math subtasks 5b, capturing reading comprehension and the ability to perform two digit subtraction with borrowing, respectively – we estimate a positive, large, and statistically significant interaction term.

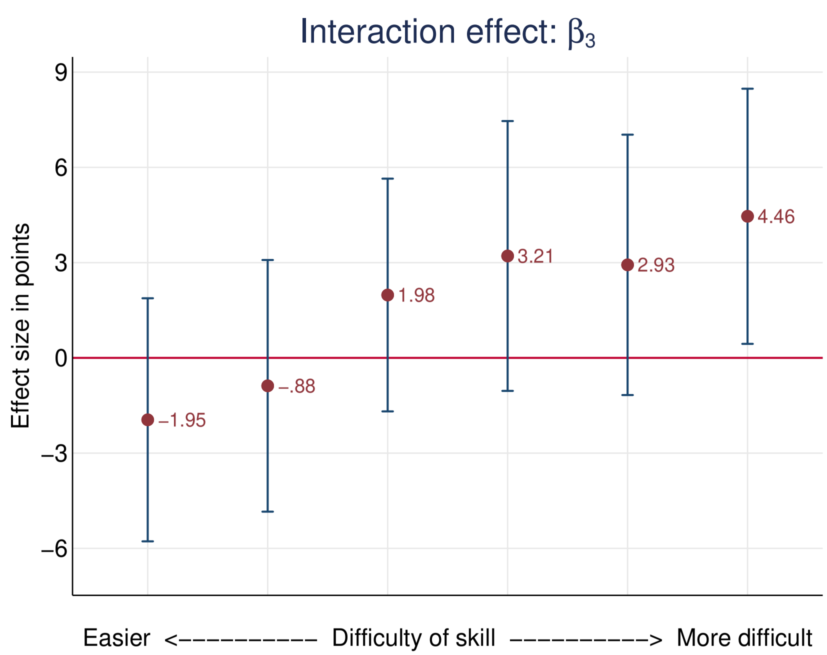
For career aspirations (Figure 4), we observe a different pattern. As with educational aspirations, we estimate a statistically significant negative estimate of β_3 for the lowest-level skills, which indicates substitutability between career aspirations and educational supply in the production of these skills. As skill difficulty increases, however, we find far less evidence of complementarity – i.e., positive estimates – between career aspirations and educational supply.[25]

To complement these results, we present full regression output in Tables A.4 and A.5. These show the point estimates and standard errors presented in Figures 3 and 4, as well as the parameter we call the interaction mean (described in Section 3) and a p-value of a test of the null that it is not different from zero. The interaction mean provides a test for whether, in intervention villages, children of high-aspirations caregivers demonstrate a higher endline level of the skill in question than do children of other families in the same village. For educational aspirations, our estimates of the interaction mean are positive and statistically significant for both reading and math for all but the lowest-level subtasks. For career aspirations, our estimates of this parameter are much smaller in magnitude and never statistically significant.

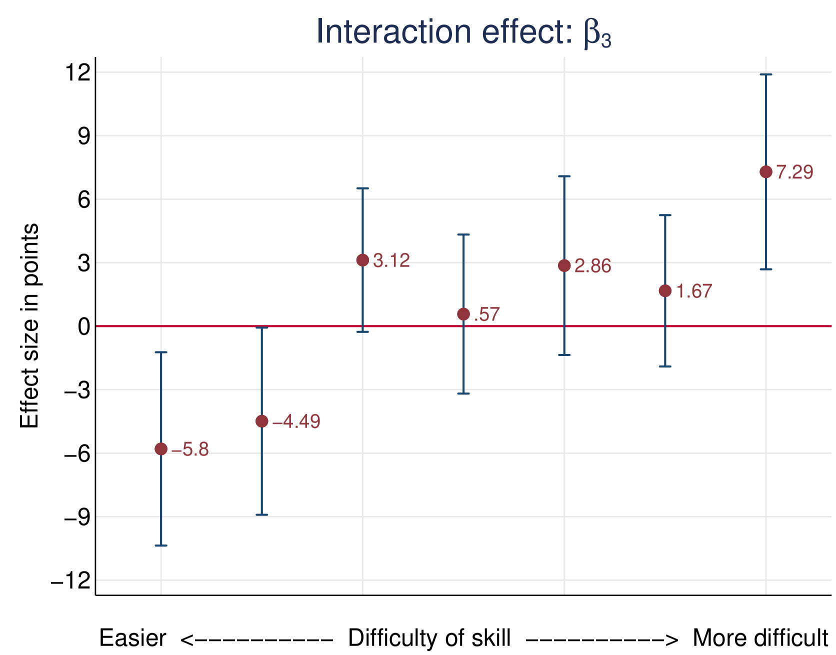
Together, these results show how demand and supply interact to generate learning in this setting. For the easiest reading and math skills, our estimates suggest substitutability between supply and demand, as measured by both educational and career aspirations. For higher-level subtasks which are closer to literacy and numeracy, we observe complementarity between educational sup-

[25] Subtask 6 on both tests has no written component, making it somewhat different than all other subtasks, and less difficult in practice than other higher-level subtasks.

Figure 3: How educational aspirations and educational supply combine to produce reading and math skills



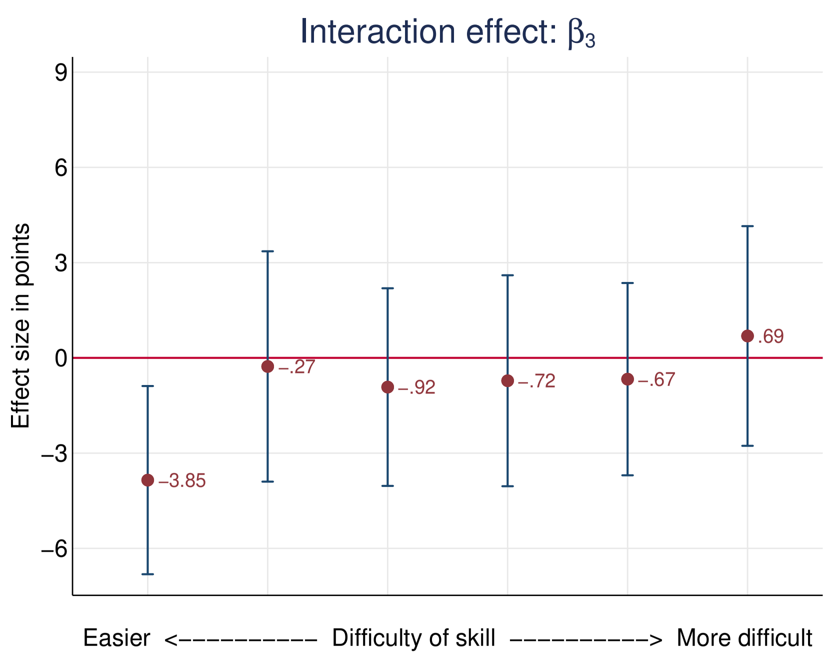
Panel A: Reading skills



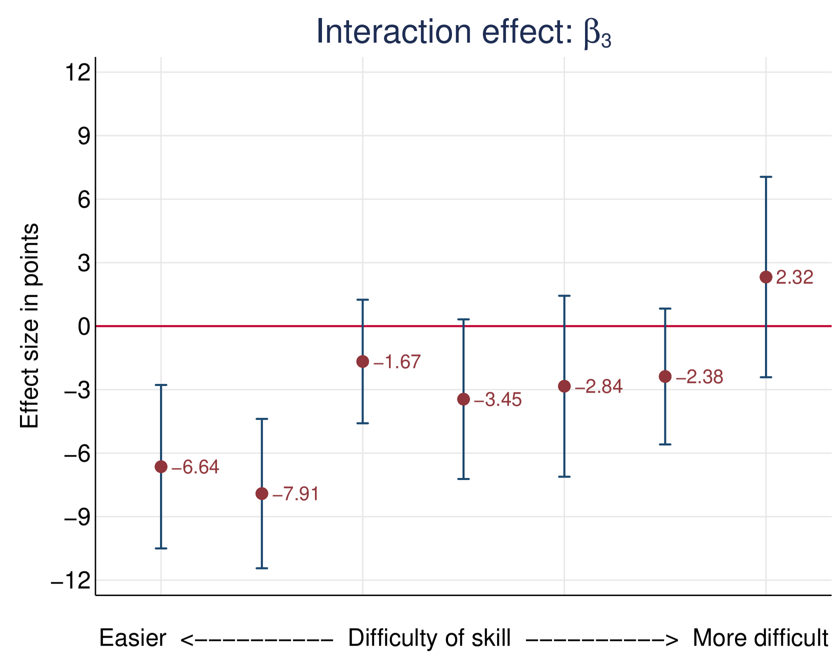
Panel B: Math skills

Figure 3 notes: this figure shows point estimates of the parameter β_3 and their confidence intervals, estimated for the interaction between baseline educational aspirations and the intervention-generated increase in the quality of educational supply. We present separate estimates for performance on each subtask – for reading (Panel A) and math (Panel B), respectively – as the dependent variable. The subtasks for reading are 1, 2, 3, 4, 5a, and 5b. For math, they are 1, 2, 3, 4a, 4b, 5a, 5b. We relabel the x-axis and suppress subtask labels there for clarity. We omit subtask 6 in each graph because it does not contain a written component (see Footnote 25 and Banerjee et al., 2017b). All 3,813 observations in our estimation sample from Table 6 were used to generate these figures.

Figure 4: How career aspirations and educational supply combine to produce reading and math skills



Panel A: Reading skills



Panel B: Math skills

Figure 4 notes: this figure shows point estimates of the parameter β_3 and their confidence intervals, estimated for the interaction between baseline career aspirations and the intervention-generated increase in the quality of educational supply. We present separate estimates for performance on each subtask – for reading (Panel A) and math (Panel B), respectively – as the dependent variable. The subtasks for reading are 1, 2, 3, 4, 5a, and 5b. For math, they are 1, 2, 3, 4a, 4b, 5a, 5b. We relabel the x-axis and suppress subtask labels there for clarity. We omit subtask 6 in each graph because it does not contain a written component (see Footnote 25 and Banerjee et al., 2017b). All 3,813 observations in our estimation sample from Table 6 were used to generate these figures.

ply and educational aspirations, but no evidence of such a relationship for career aspirations.

We believe this shows that the dramatic increase in the quality of educational supply shifted the impact of the marginal unit of family support – be it encouragement or investment – that high-aspirations families make. In the status quo group, we estimate the largest differences between the children of low- and high-aspirations caregivers in their performance on the lowest level subtasks (see Figure 2 and Tables A.2 and A.3). For children in the intervention group whose caregivers express high educational aspirations, our results in this section show that these differences at lower levels disappear, but at higher levels are now far larger than in the status quo group. The fact that we do not see this pattern for higher-level subtasks among children of caregivers with high career aspirations could suggest that families with high career aspirations take an approach to their child's learning that is closer to satisficing. It is also consistent with the notion that career aspirations differ from educational aspirations in terms of how they are acted upon, as we saw previously for our analysis of educational investment (i.e., in column 2 of Table 4).

5.3 Alternative explanations

In this section we address two potential alternative explanations for our results on the interaction between supply and demand. The first is that they reflect the correlation between unobserved child ability and aspirations. The second is that they reflect the correlation between unobserved family wealth and aspirations.

It is difficult to infer the likely sign of effects from other contributing sources here using the type of bounding exercise used in Section 4.4. This is because the intervention could be either a substitute or complement for inputs such as household wealth or unobserved child ability. If these inputs were complements, our estimates would be an upper bound, as the true mapping from aspirations and the intervention would be smaller. If they were substitutes, our estimates would likely be a lower bound. In this section we investigate the likelihood of, and empirical evidence for influence from such unobserved contributors.

For several reasons, unobserved child ability – and its correlation with aspirations – is highly unlikely to be the main explanation for our results. First, caregivers are highly unlikely to know whether the child is of high academic ability at the time that we measure baseline aspirations. These data were collected when the child had not yet been to school, so the family would have received no feedback from teachers. Furthermore, as Dizon-Ross (2019) documents, even after children enroll in school, caregivers in low-income settings often have highly inaccurate beliefs about child ability.[26] In addition, caregivers are unlikely to be able to assess academic ability through the lens of their own academic experience, since more than three quarters of the caregivers in our sample have never been to school themselves, and over 90 percent of them could not read a short, simple sentence at baseline.

Second, the mapping from aspirations to educational investment we measure is similar to that found in another context. In rural Ethiopia, Bernard et al. (2014) report a statistically significant increase in educational investment in response to an experimentally-generated increase in aspirations. Their estimate of this relationship is very similar in magnitude to ours. Third, while career and education aspirations both predict subsequent investment behavior, they are only mildly correlated (pairwise correlation: 0.18).

We can also examine how much aspirations vary across children, within a family; this tests for unobservable, within-family differences in child ability manifesting via aspirations. There are 151 caregivers in our sample with more than one child who is enrolled in our study. In 92 percent of these cases, the caregiver expresses the same educational aspirations for each child under their care. In 70 percent of these cases, the caregiver expresses the same career aspirations for each child under their care. This suggests that our measures of aspirations capture family desire for the future of (all of) their children, rather than family beliefs about an individual child's skill or ability.

There are also several reasons why it is highly unlikely that some broader, latent socioeconomic variable drives our estimates of the interaction between baseline aspirations and educational supply. We first conduct a robustness test, estimating an alternative version of Table 6 with interactions

[26]Gallegos and Celhayb (Forthcoming) show that in a much higher-income context, Chile, parent beliefs respond to signals from the school about child ability, and that this process occurs over several years after the child first enters school.

between the intervention and household wealth, caregiver education, caregiver literacy, and the presence of books in the home. In Tables A.6 and A.7 we present results for baseline educational and career aspirations, respectively. These show that the main patterns we observe in Table 6 are robust to the inclusion of these other predictors of a potential non-aspirations response to the intervention. In other words, for a reasonable set of observable controls, we show that there is a residual in the learning outcomes that we study. This residual is not explained by the interaction of the intervention and these other traits of the children and their families which also predict learning, but it can partly be explained by differentials in baseline educational aspirations.

We present a similar set of results, only interacting the intervention and these same traits, in Tables A.8 and A.9. These show that the household characteristics we control for that are associated with socio-economic status – household wealth, whether the caregiver has been to school, and caregiver literacy – do not have statistically significant positive interactions with the intervention in predicting the outcome variable. Caregiver literacy, interacted with the treatment, has a negative and statistically significant estimated relationship with the endline test score, but a positive estimated relationship with literacy similar in magnitude to the relationship between educational aspirations interacted with the intervention and learning. Its relationship with numeracy, however, is negative and much smaller. This shows further evidence of interaction between household traits and the quality of education supplied in the generation of learning in such areas.

5.4 Aspirations failure and systems failure

In the active literature on the economics of aspirations, there are two key links: one, from aspirations to actions, usually investment in education, business, or some other endeavor with potentially high future returns; and two, from these actions to outcomes, usually educational attainment, learning, or enterprise profits. Dalton et al. (2016) builds a theoretical model in which people can hold suboptimally high aspirations, such that if there exists an insurmountably large gap between the aspiration and the person's current state, the person may choose to invest very little. They refer to this state as "aspirations frustration" or "aspirations failure." Ross (2019) shows empirical evidence of this phenomenon in educational investment in rural India, and McKenzie et al. (2022) show evidence of it among entrepreneurs in the Philippines.[27]

Seen through this lens, our results show that there can be systems failure even when there is no aspirations failure. We show that in status quo villages, the first link is intact: higher baseline aspirations map onto to significantly greater subsequent investment – i.e., no evidence of aspirations failure. Nonetheless, the second link breaks down: these investments yield zero or very little gain in terms of the acquisition of foundational literacy and numeracy skills.

We then show that this breakdown of the second link, between actions and outcomes, does not have to be the case. With the benefit of high-quality educational supply, high-demand families are able to help their children master key higher-level skills beyond what children of other families in these same villages achieve. This suggests that the failure here is of the system, i.e., a "systems failure," in juxtaposition to the aspirations failure or aspirations frustration studied elsewhere.

6 Conclusion

Across the world, many families wish for their children to live better lives than those lived by previous generations, and a common path for realizing this desire is through education. We characterize this process in a context of extreme poverty, estimating how family inputs and school system inputs interact to generate learning, via the educational system, in a crucial stage of early childhood.

We show that high-demand caregivers expend greater amounts of dear household resources to invest in their children's schooling, both in terms of money and their children's time. These investments yield a positive return in terms of the child's relative performance on literacy and numeracy tests, with children of these caregivers performing roughly 0.3 SD better than other children on endline tests.

Sadly, because counterfactual learning levels are extremely low in the rural Gambian status quo, these relative gains still leave children nowhere near achieving developmentally meaningful

[27] Leight et al. (2021) report the evaluation of an intervention to raise aspirations in Ethiopia, similar to but distinct from that studied in Bernard et al. (2014); they find no measurable effect of the intervention on either aspirations or investment.

levels of learning, particularly literacy or numeracy. These are among the most crucial skills for reaching later economic productivity and participating in many spheres of society.

With the presence of complementary inputs on the supply side, however, we show that family demand does map onto far greater likelihood of the child mastering developmentally meaningful skills, including the ability to read with understanding and conduct basic arithmetic. For research, this suggests the need for greater study of how demand and supply interact to create learning at different levels of economic prosperity. For policy, this suggests that while the demand side can yield important learning gains in some low- and middle-income settings, substantial increases in the quality of educational supply will also be necessary to address the very low levels of learning in the many pockets of extreme poverty in the developing world.

References

Alesina, Alberto, Sebastian Hohmann, Stelios Michalopoulos, and Elias Papaioannou, "Intergenerational mobility in Africa," *Econometrica*, 2021, *89* (1), 1–35.

Asadullah, M Niaz, Md Abdul Alim, and M Anowar Hossain, "Enrolling girls without learning: Evidence from public schools in Afghanistan," *Development Policy Review*, 2019, *37* (4), 486–503.

Asher, Sam, Paul Novosad, and Charlie Rafkin, "Intergenerational mobility in India: Estimates from new methods and administrative data," *World Bank Working Paper*, 2018.

Attanasio, Orazio, Costas Meghir, and Emily Nix, "Human capital development and parental investment in India," *Review of Economic Studies*, 2020, *87* (6), 2511–2541.

Azam, Mehtabul and Vipul Bhatt, "Like father, like son? Intergenerational educational mobility in India," *Demography*, 2015, *52* (6), 1929–1959.

Banerjee, Abhijit, Rukmini Banerji, James Berry, Esther Duflo, Harini Kannan, Shobhini Mukerji, Marc Shotland, and Michael Walton, "From proof of concept to scalable policies: Challenges and solutions, with an application," *Journal of Economic Perspectives*, 2017, *31* (4), 73–102.

Banerjee, Abhijit V, Swati Bhattacharjee, Raghabendra Chattopadhyay, and Alejandro J Ganimian, "The untapped math skills of working children in India: Evidence, possible explanations, and implications," *Unpublished manuscript*, 2017.

Beaman, Lori, Esther Duflo, Rohini Pande, and Petia Topalova, "Female leadership raises aspirations and educational attainment for girls: A policy experiment in India," *Science*, 2012, *335*, 582–586.

Behrman, Jere R, "Investment in education–inputs and incentives," *Handbook of Development Economics*, 2010, *5*, 4883–4975.

Bernard, Tanguy, Stefan Dercon, Kate Orkin, and Alemayehu Taffesse, "The future in mind: Aspirations and forward-looking behaviour in rural Ethiopia," *BREAD Working Paper No. 429*, 2014.

Black, Sandra E, Paul J Devereux et al., "Recent developments in intergenerational mobility," *Handbook of Labor Economics*, 2011, *4*, 1487–1541.

Bommier, Antoine and Sylvie Lambert, "Education demand and age at school enrollment in Tanzania," *Journal of Human Resources*, 2000, pp. 177–203.

Boone, Peter, Alex Eble, Diana Elbourne, Samory Fernandes, Chris Frost, Chitra Jayanty, Maitri Lenin, Ana Filipa Silva et al., "Remedial after-school support classes offered in rural Gambia (The SCORE trial): study protocol for a cluster randomized controlled trial," *Trials*, 2015, *16* (1), 1–9.

Bruhn, Miriam and David McKenzie, "In pursuit of balance: Randomization in practice in development field experiments," *American Economic Journal: Applied Economics*, 2009, *1* (4), 200–232.

Chetty, Raj, David Grusky, Maximilian Hell, Nathaniel Hendren, Robert Manduca, and Jimmy Narang, "The fading American dream: Trends in absolute income mobility since 1940," *Science*, 2017, *356* (6336), 398–406.

_ , Nathaniel Hendren, Patrick Kline, Emmanuel Saez, and Nicholas Turner, "Is the United States still a land of opportunity? Recent trends in intergenerational mobility," *American Economic Review*, 2014, *104* (5), 141–47.

CIA, "The world factbook 2019," *Central Intelligence Agency, Washington, DC*, 2019.

Cohodes, Sarah R and Katharine S Parham, "Charter Schools' Effectiveness, Mechanisms, and Competitive Influence," in "Oxford Research Encyclopedia of Economics and Finance" 2021.

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

Dalton, Patricio S, Sayantan Ghosal, and Anandi Mani, "Poverty and aspirations failure," *The Economic Journal*, 2016, *126* (590), 165–188.

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

Dubeck, Margaret M and Amber Gove, "The early grade reading assessment (EGRA): Its theoretical foundation, purpose, and limitations," *International Journal of Educational Development*, 2015, *40*, 315–322.

Duncan, Greg J, Chantelle J Dowsett, Amy Claessens, Katherine Magnuson, Aletha C Huston, Pamela Klebanov, Linda S Pagani, Leon Feinstein, Mimi Engel, Jeanne Brooks-Gunn et al., "School readiness and later achievement.," *Developmental Psychology*, 2007, *43* (6), 1428.

Eble, Alex, Chris Frost, Alpha Camara, Baboucarr Bouy, Momodou Bah, Maitri Sivaraman, Jenny Hsieh, Chitra Jayanty, Tony Brady, Piotr Gawron, Peter Boone, and Diana Elbourne, "How much can we remedy very low learning levels in rural parts of low-income countries? Impact and generalizability of a multi-pronged para-teacher intervention from a cluster-randomized trial in The Gambia," *Journal of Development Economics*, 2021, *148* (102539).

Evans, David K and Fei Yuan, "How Big Are Effect Sizes in International Education Studies?," *Educational Evaluation and Policy Analysis*, 2022, *44* (3), 532–540.

Fazzio, Ila, Alex Eble, Robin L Lumsdaine, Peter Boone, Baboucarr Bouy, Pei-Tseng Jenny Hsieh, Chitra Jayanty, Simon Johnson, and Ana Filipa Silva, "Large Learning Gains in Pockets of Extreme Poverty: Experimental Evidence from Guinea Bissau," *Journal of Public Economics*, 2021, *199* (104385).

Filmer, Deon, Halsey Rogers, Noam Angrist, and Shwetlena Sabarwal, "Learning-adjusted years of schooling (LAYS): Defining a new macro measure of education," *Economics of Education Review*, 2020, *77*, 101971.

Foster, Andrew D and Mark R Rosenzweig, "Technical change and human-capital returns and investments: evidence from the green revolution," *American Economic Review*, 1996, pp. 931–953.

Fruttero, Anna, Noel Muller, and Oscar Calvo-Gonzalez, "The power and roots of aspirations: A survey of the empirical evidence," *World Bank Policy Research Working Paper Number 9729*, 2021.

Furr, R Michael, *Psychometrics: an introduction*, SAGE publications, 2021.

Gallegos, Sebastian and Pablo A Celhayb, "Early skill effects on types of parental investments and long-run outcomes," *Journal of Human Resources*, Forthcoming.

Ganimian, Alejandro J and Richard J Murnane, "Improving education in developing countries: Lessons from rigorous impact evaluations," *Review of Educational Research*, 2016, *86* (3), 719–755.

Genicot, Garance and Debraj Ray, "Aspirations and inequality," *Econometrica*, 2017, *85* (2), 489–519.

Glewwe, Paul and Hanan G Jacoby, "An economic analysis of delayed primary school enrollment in a low income country: the role of early childhood nutrition," *Review of Economics and Statistics*, 1995, pp. 156–169.

_ **and Karthik Muralidharan**, "Improving education outcomes in developing countries: Evidence, knowledge gaps, and policy implications," in "Handbook of the Economics of Education," Vol. 5 2016, pp. 653–743.

Hill, Carolyn J, Howard S Bloom, Alison Rebeck Black, and Mark W Lipsey, "Empirical benchmarks for interpreting effect sizes in research," *Child development perspectives*, 2008, *2* (3), 172–177.

Jackson, C Kirabo, Rucker C Johnson, and Claudia Persico, "The effects of school spending on educational and economic outcomes: Evidence from school finance reforms," *Quarterly Journal of Economics*, 2016, *131* (1), 157–218.

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

Kerwin, Jason T and Rebecca L Thornton, "Making the grade: The sensitivity of education program effectiveness to input choices and outcome measures," *Review of Economics and Statistics*, 2021, *103* (2), 251–264.

Kingdon, Geeta Gandhi and Vandana Sipahimalani-Rao, "Para-teachers in India: Status and impact," *Economic and Political weekly*, 2010, pp. 59–67.

Kipchumba, Elijah Kipkech, Catherine Porter, Danila Serra, Munshi Sulaiman et al., "Infuencing youths' aspirations and gender attitudes through role models: Evidence from Somali schools," *Working Paper*, 2021.

Knudsen, Eric I, "Sensitive periods in the development of the brain and behavior," *Journal of Cognitive Neuroscience*, 2004, *16* (8), 1412–1425.

Kraft, Matthew A, "Interpreting effect sizes of education interventions," *Educational Researcher*, 2020, *49* (4), 241–253.

_, David Blazar, and Dylan Hogan, "The effect of teacher coaching on instruction and achievement: A meta-analysis of the causal evidence," *Review of Educational Research*, 2018, *88* (4), 547–588.

Kremer, Michael and Alaka Holla, "Improving education in the developing world: What have we learned from randomized evaluations?," *Annual Review of Economics*, 2009, *1* (1), 513–542.

Krueger, Alan B and Diane M Whitmore, "The effect of attending a small class in the early grades on college-test taking and middle school test results: Evidence from Project STAR," *The Economic Journal*, 2001, *111* (468), 1–28.

La Ferrara, Eliana, "Aspirations, social norms, and development," *Journal of the European Economic Association*, 2019, *17* (6), 1687–1722.

Leight, Jessica, Daniel Gilligan, Michael Mulford, Alemayehu Seyoum Taffesse, and Heleene Tambet, "Aspiring to more? New evidence on the effect of light-touch aspirations interventions in rural Ethiopia," *Working Paper*, 2021.

Lin, Winston and Donald P Green, "Standard operating procedures: A safety net for pre-analysis plans," *PS: Political Science & Politics*, 2016, *49* (3), 495–500.

Lybbert, Travis J and Bruce Wydick, "Poverty, aspirations, and the economics of hope," *Economic Development and Cultural Change*, 2018, *66* (4), 709–753.

Mbiti, Isaac, Karthik Muralidharan, Mauricio Romero, Youdi Schipper, Constantine Manda, and Rakesh Rajani, "Inputs, incentives, and complementarities in education: Experimental evidence from Tanzania," *Quarterly Journal of Economics*, 2019, *134* (3), 1627–1673.

McEwan, Patrick J, "Improving learning in primary schools of developing countries: A meta-analysis of randomized experiments," *Review of Educational Research*, 2014.

McKenzie, David, Aakash Mohpal, and Dean Yang, "Aspirations and financial decisions: Experimental evidence from the Philippines," *Journal of Development Economics*, 2022, *156*, 102846.

Muralidharan, Karthik, Abhijeet Singh, and Alejandro J Ganimian, "Disrupting education? Experimental evidence on technology-aided instruction in India," *American Economic Review*, 2019, *109* (4), 1426–60.

— **and Venkatesh Sundararaman**, "Contract teachers: Experimental evidence from India," *NBER Working Paper 19440*, 2013.

Nelson III, Charles A and Laurel J Gabard-Durnam, "Early adversity and critical periods: neurodevelopmental consequences of violating the expectable environment," *Trends in Neurosciences*, 2020, *43* (3), 133–143.

Nonoyama-Tarumi, Yuko, Edilberto Loaiza, and Patrice L Engle, "Late entry into primary school in developing societies: Findings from cross-national household surveys," *International Review of Education*, 2010, *56* (1), 103–125.

Olken, Benjamin A, "Promises and perils of pre-analysis plans," *Journal of Economic Perspectives*, 2015, *29* (3), 61–80.

Piper, Benjamin, Joseph Destefano, Esther M Kinyanjui, and Salome Ong'ele, "Scaling up successfully: Lessons from Kenya's Tusome national literacy program," *Journal of Educational Change*, 2018, *19* (3), 293–321.

— **, Stephanie Simmons Zuilkowski, and Abel Mugenda**, "Improving reading outcomes in Kenya: First-year effects of the PRIMR Initiative," *International Journal of Educational Development*, 2014, *37*, 11–21.

Platas, LM, L Ketterlin-Gellar, A Brombacher, and Y Sitabkhan, "Early grade mathematics assessment (EGMA) toolkit," *Research Triangle Park, NC: RTI International*, 2014.

Pratham, New Delhi, "Annual status of education report 2012," 2012.

Pritchett, Lant, *The rebirth of education: Schooling ain't learning*, CGD Books, 2013.

— **and Amanda Beatty**, "Slow down, you're going too fast: Matching curricula to student skill levels," *International Journal of Educational Development*, 2015, *40*, 276–288.

Romero, Mauricio, Justin Sandefur, and Wayne Aaron Sandholtz, "Outsourcing education: Experimental evidence from Liberia," *American Economic Review*, 2020, *110* (2), 364–400.

Ross, Phillip H, "Occupation aspirations, education investment, and cognitive outcomes: Evidence from Indian adolescents," *World Development*, 2019, *123*, 104613.

Serneels, Pieter and Stefan Dercon, "Aspirations, poverty, and education: Evidence from India," *The Journal of Development Studies*, 2021, *57* (1), 163–183.

Sprenger-Charolles, Liliane, "The Gambia Early Grade Reading Assessment: Results from 1,200 Gambian primary students learning to read in English," *World Bank Policy Report*, 2008, (69716).

The Gambia Bureau of Statistics and ICF International, "The Gambia Demographic and Health Survey 2013," 2014.

Walters, Christopher R, "The demand for effective charter schools," *Journal of Political Economy*, 2018, *126* (6), 2179–2223.

Werker, Janet F and Richard C Tees, "Speech perception as a window for understanding plasticity and commitment in language systems of the brain," *Developmental Psychobiology: The Journal of the International Society for Developmental Psychobiology*, 2005, *46* (3), 233–251.

Wolf, Sharon and Dana Charles McCoy, "The role of executive function and social-emotional skills in the development of literacy and numeracy during preschool: a cross-lagged longitudinal study," *Developmental Science*, 2019, *22* (4), e12800.

Wooldridge, Jeffrey M, *Introductory econometrics: A modern approach*, Nelson Education, 2016.

Wright, Donald R, *The world and a very small place in Africa: a history of globalization in Niumi, the Gambia*, Routledge, 2015.

Appendix

Table A.1: Estimating the mapping of aspirations at baseline to subsequent educational investment, including both status quo and intervention groups

	(1) Educational expenditure	(2) School-related time use	(3) Enrolled in school, year 1	(4) Enrolled in school, year 2	(5) Enrolled in school, year 3
Panel A: Educational aspirations					
Aspirations x intervention (β_3)	4.88**	-0.017*	0.071*	-0.003**	0.005
	(40.26)	(0.009)	(0.042)	(0.035)	(0.011)
Intervention (β_2)	-79.23*	0.139***	-0.069	0.040	0.002
	(39.10)	(0.011)	(0.043)	(0.033)	(0.012)
Aspirations (β_1)	79.46**	0.021***	0.034	0.059**	0.008
	(28.69)	(0.007)	(0.026)	(0.026)	(0.007)
Comparison group mean	572.54	0.611	0.794	0.822	0.972
Number of observations	3,654	3,732	3,754	3,702	3,732
Panel B: Career aspirations					
Aspirations x intervention (β_3)	4.33**	-0.001	0.035	-0.010***	-0.000
	(40.71)	(0.010)	(0.034)	(0.029)	(0.008)
Intervention (β_2)	-79.20**	0.129***	-0.048	0.045	0.006
	(38.14)	(0.012)	(0.042)	(0.033)	(0.010)
Aspirations (β_1)	66.38**	0.008	0.035	0.058***	0.002
	(27.72)	(0.006)	(0.023)	(0.020)	(0.005)
Comparison group mean	576.20	0.614	0.802	0.823	0.977
Number of observations	3,654	3,732	3,754	3,702	3,732

Table A.1 notes: this presents an analog to Table 4, but including children from both the status quo and intervention groups. Here we report the results of estimating Equation 2 using the outcome variable given in the column heading and the baseline aspirations (educational or career) indicated in the panel heading, including both the status quo and intervention groups. Dependent variables labeled in the column headings are defined in the text. We report clustered standard errors in parentheses below each estimated coefficient. Observations vary by column because outcome variables were collected at different times and some children were missed in some periods. Results are robust to including only the smallest estimation sample. The full set of controls is as indicated in Section 3. $^{*}p < 0.10$, $^{**}p < 0.05$, $^{***}p < 0.01$.

Table A.2: Mapping of aspirations at baseline to endline performance on reading subtasks in the status quo group

	Subtask 1	Subtask 2	Subtask 3	Subtask 4	Subtask 5a	Subtask 5b	Subtask 6
Panel A: Educational aspirations							
High baseline educational aspirations (α_1)	3.364**	3.740***	1.635*	2.395**	2.291**	1.212**	1.126
	(1.327)	(1.294)	(0.898)	(0.904)	(0.925)	(0.534)	(0.889)
Comparison group mean	11.592	25.741	3.744	3.729	4.371	2.028	4.309
Number of observations	2,039	2,039	2,039	2,039	2,039	2,039	2,039
Panel B: Career aspirations							
High baseline career aspirations (α_1)	3.955***	2.769**	2.129***	2.400***	2.202***	1.635***	2.256***
	(0.855)	(1.210)	(0.511)	(0.603)	(0.557)	(0.308)	(0.641)
Comparison group mean	10.884	25.949	3.295	3.499	4.183	1.671	3.494
Number of observations	2,039	2,039	2,039	2,039	2,039	2,039	2,039

Table A.2 notes: this table shows results for estimating Equation 1 for children's scores on the individual reading subtasks. We restrict our attention in this table to children in the status quo group. We report clustered standard errors in parentheses below each estimated coefficient. The dependent variable in each column is the subtask number listed in the column heading. Subtasks are described in Table 2. The tests are shown in their entirety in Appendix A, divided by subtasks. Each subtask number is indicated at the top of each relevant block of questions. The possible values of each subtask score range from zero to 100 percent of questions answered correctly. $^{*}p < 0.10$, $^{**}p < 0.05$, $^{***}p < 0.01$.

Table A.3: Mapping of aspirations at baseline to endline performance on math subtasks in the status quo group

	Subtask 1	Subtask 2	Subtask 3	Subtask 4a	Subtask 4b	Subtask 5a	Subtask 5b	Subtask 6
Panel A: Educational aspirations								
High baseline educational aspirations (α_1)	7.286***	7.716***	2.229**	3.138***	2.740***	2.695***	0.734	4.766***
	(2.053)	(1.840)	(0.982)	(1.127)	(0.798)	(0.876)	(0.657)	(1.135)
Comparison group mean	41.153	33.866	16.109	14.594	6.337	7.414	2.978	21.779
Number of observations	2,038	2,038	2,038	2,038	2,038	2,038	2,038	2,038
Panel B: Career aspirations								
High baseline career aspirations (α_1)	6.642***	8.287***	2.820***	4.879***	4.474***	2.989***	1.166**	3.412***
	(1.662)	(1.459)	(0.744)	(0.978)	(0.782)	(0.622)	(0.572)	(0.941)
Comparison group mean	41.183	33.074	15.623	13.371	4.958	7.132	2.597	22.450
Number of observations	2,038	2,038	2,038	2,038	2,038	2,038	2,038	2,038

Table A.3 notes: this table shows results for estimating Equation 1 for children's scores on the individual math subtasks. We restrict our attention in this table to children in the status quo group. We report clustered standard errors in parentheses below each estimated coefficient. The dependent variable in each column is the subtask number listed in the column heading. Subtasks are described in Table 2. The tests are shown in their entirety in Appendix A, divided by subtasks. Each subtask number is indicated at the top of each relevant block of questions. The possible values of each subtask score range from zero to 100 percent of questions answered correctly. $^{*}p < 0.10$, $^{**}p < 0.05$, $^{***}p < 0.01$.

Table A.4: Demand, supply, and math skill acquisition

	Subtask 1	Subtask 2	Subtask 3	Subtask 4a	Subtask 4b	Subtask 5a	Subtask 5b	Subtask 6
Panel A: Educational aspirations								
Aspirations x intervention (β_3)	-5.805**	-4.490**	3.125*	0.574	2.868	1.678	7.296***	0.141
	(2.330)	(2.256)	(1.730)	(1.917)	(2.154)	(1.824)	(2.349)	(1.956)
Intervention (β_2)	50.171***	50.003***	41.133***	46.512***	56.478***	38.885***	46.813***	26.500***
	(2.940)	(2.796)	(1.853)	(2.170)	(2.207)	(1.635)	(2.259)	(1.966)
Aspirations (β_1)	7.940***	8.351***	2.399**	3.523***	2.952***	3.013***	0.906	5.078***
	(2.047)	(1.841)	(0.957)	(1.102)	(0.816)	(0.842)	(0.681)	(1.119)
Interaction mean (β_1 + β_3)	2.135	3.861	5.524	4.097	5.820	4.691	8.202	5.219
P-value [β_1 + β_3 = 0]	[0.053]	[0.003]	[0.000]	[0.012]	[0.004]	[0.005]	[0.000]	[0.001]
Comparison group mean	49.916	42.964	19.092	18.574	9.469	10.741	3.971	27.198
Number of observations	3,813	3,813	3,813	3,813	3,813	3,813	3,813	3,813
Panel B: Career aspirations								
Aspirations x intervention (β_3)	-6.645***	-7.916***	-1.671	-3.450*	-2.844	-2.382	2.324	-2.295
	(1.970)	(1.800)	(1.489)	(1.924)	(2.181)	(1.637)	(2.415)	(1.930)
Intervention (β_2)	50.925***	52.385***	44.103***	49.091***	60.058***	41.435***	49.719***	28.041***
	(2.729)	(2.566)	(1.903)	(2.145)	(2.450)	(1.731)	(2.590)	(2.031)
Aspirations (β_1)	7.318***	9.010***	2.972***	5.282***	4.717***	3.330***	1.272**	3.662***
	(1.668)	(1.472)	(0.735)	(0.991)	(0.830)	(0.639)	(0.581)	(0.943)
Interaction mean (β_1 + β_3)	0.673	1.094	1.301	1.832	1.873	0.948	3.596	1.367
P-value [β_1 + β_3 = 0]	[0.520]	[0.299]	[0.312]	[0.264]	[0.349]	[0.527]	[0.127]	[0.416]
Comparison group mean	49.357	42.818	19.163	18.972	10.002	10.684	4.111	26.507
Number of observations	3,813	3,813	3,813	3,813	3,813	3,813	3,813	3,813

Table A.4 notes: this table shows results for estimating Equation 2 for children's scores on the individual math subtasks; panel titles indicate which aspirations are being studied. The dependent variable in each column is the subtask listed in the column heading; subtasks are described in Table 2. We report clustered standard errors in parentheses below each estimated coefficient. The tests are shown in their entirety in Appendix A, divided by subtasks. Each subtask number is indicated at the top of each relevant block of questions. The possible values of each subtask score range from zero to 100 percent of questions answered correctly. $^{*}p < 0.10$, $^{**}p < 0.05$, $^{***}p < 0.01$.

Table A.5: Demand, supply, and reading skill acquisition

	Subtask 1	Subtask 2	Subtask 3	Subtask 4	Subtask 5a	Subtask 5b	Subtask 6
Panel A: Educational aspirations							
Aspirations x intervention (β_3)	-1.955	-0.882	1.989	3.215	2.938	4.464**	-1.039
	(1.953)	(2.022)	(1.870)	(2.168)	(2.092)	(2.050)	(2.183)
Intervention (β_2)	55.870***	24.492***	45.678***	57.489***	54.465***	41.945***	57.149***
	(2.161)	(2.111)	(1.907)	(2.243)	(2.283)	(2.097)	(2.373)
Aspirations (β_1)	3.559***	4.136***	1.767**	2.425***	2.462***	1.193**	1.083
	(1.302)	(1.294)	(0.883)	(0.868)	(0.893)	(0.507)	(0.861)
Interaction mean (β_1 + β_3)	1.604	3.254	3.756	5.640	5.400	5.657	0.044
P-value [β_1 + β_3 = 0]	[0.275]	[0.042]	[0.026]	[0.005]	[0.005]	[0.006]	[0.983]
Comparison group mean	15.444	29.952	5.710	6.447	7.060	3.360	5.918
Number of observations	3,814	3,814	3,814	3,814	3,814	3,814	3,814
Panel B: Career aspirations							
Aspirations x intervention (β_3)	-3.853**	-0.278	-0.927	-0.721	-0.672	0.699	-1.601
	(1.512)	(1.851)	(1.588)	(1.695)	(1.545)	(1.765)	(2.280)
Intervention (β_2)	57.172***	24.108***	47.479***	59.892***	56.667***	44.191***	57.560***
	(2.018)	(2.094)	(1.956)	(2.065)	(2.053)	(2.067)	(2.395)
Aspirations (β_1)	4.160***	3.125**	2.203***	2.320***	2.253***	1.586***	2.389***
	(0.873)	(1.211)	(0.527)	(0.605)	(0.573)	(0.349)	(0.666)
Interaction mean (β_1 + β_3)	0.307	2.847	1.276	1.599	1.581	2.285	0.788
P-value [β_1 + β_3 = 0]	[0.802]	[0.034]	[0.390]	[0.310]	[0.269]	[0.187]	[0.713]
Comparison group mean	15.579	29.582	5.825	6.400	6.992	3.465	6.248
Number of observations	3,814	3,814	3,814	3,814	3,814	3,814	3,814

Table A.5 notes: this table shows results for estimating Equation 2 for children's scores on the individual reading subtasks; panel titles indicate which aspiration is being studied. The dependent variable in each column is the subtask listed in the column heading; subtasks are described in Table 2. We report clustered standard errors in parentheses below each estimated coefficient. The tests are shown in their entirety in Appendix A, divided by subtasks. Each subtask number is indicated at the top of each relevant block of questions. The possible values of each subtask score range from zero to 100 percent of questions answered correctly. $^{*}p < 0.10$, $^{**}p < 0.05$, $^{***}p < 0.01$.

Table A.6: How the mapping from baseline aspirations to endline learning changes with a large increase in the quality of educational supply, including interactions with various other predictors of learning

	(1) Endline test score	(2) Child is literate	(3) Child is numerate	(4) Words read per minute
Educational aspirations x intervention (β_3)	0.32	0.06***	0.04*	3.07*
	(1.53)	(0.02)	(0.02)	(1.58)
Educational aspirations x household wealth	2.30	0.01	0.02	0.99
	(1.71)	(0.03)	(0.03)	(1.45)
Educational aspirations x caregiver has never been to school	-0.00	-0.05*	-0.04**	-0.53
	(1.43)	(0.03)	(0.02)	(1.51)
Educational aspirations x caregiver can read simple sentence	-1.85	0.01	-0.05	-5.09
	(2.89)	(0.05)	(0.05)	(3.52)
Educational aspirations x books in house	3.71***	0.02	0.05***	3.17***
	(1.26)	(0.02)	(0.02)	(1.34)
Educational aspirations (β_2)	0.75	0.02	-0.00	-0.44
	(1.66)	(0.03)	(0.02)	(1.59)
Household wealth	-0.36	0.01	-0.02	0.38
	(1.31)	(0.01)	(0.02)	(0.98)
Caregiver has never been to school	-0.27	0.05**	0.02	0.95
	(1.19)	(0.02)	(0.01)	(1.41)
Caregiver can read simple sentence	5.16**	0.02	0.02	7.03**
	(2.42)	(0.04)	(0.04)	(3.25)
Books in house	-0.50	-0.02	-0.02	-1.70
	(1.23)	(0.02)	(0.02)	(1.33)
Intervention (β_1)	45.58***	0.23***	0.17***	35.29***
	(1.71)	(0.02)	(0.02)	(1.76)
Comparison group mean	14.96	0.00	0.01	1.99
Number of observations	3,814	3,814	3,813	3,805

Table A.6 notes: this table shows results for estimating Equation 2 after adding the interaction terms shown here. This is an analog to Panel A of Table 6, adding the interaction terms shown here to test whether, for a reasonable set of observable controls, there is still a residual in the learning outcomes we study to be explained by aspirations which is not explained by the interaction of the intervention and other traits of the children and their families which also predict learning. We report clustered standard errors in parentheses below each estimated coefficient. $^{*}p < 0.10$, $^{**}p < 0.05$, $^{***}p < 0.01$.

Table A.7: How the mapping from baseline aspirations to endline learning changes with a large increase in the quality of educational supply, including interactions with various other predictors of learning

	(1) Endline test score	(2) Child is literate	(3) Child is numerate	(4) Words read per minute
Career aspirations x intervention (β_3)	-2.39*	0.03	0.01	0.87
	(1.32)	(0.02)	(0.02)	(1.27)
Career aspirations x household wealth	1.65	0.00	-0.02	-0.11
	(1.48)	(0.02)	(0.02)	(1.53)
Career aspirations x caregiver has never been to school	1.77	-0.03	-0.04	1.69
	(1.95)	(0.02)	(0.03)	(1.74)
Career aspirations x caregiver can read simple sentence	3.93	0.06	0.01	5.04*
	(3.13)	(0.04)	(0.05)	(2.98)
Career aspirations x books in house	1.58	0.00	0.00	-0.27
	(1.40)	(0.02)	(0.02)	(1.51)
Career aspirations (β_1)	0.77	0.02	0.04	-0.22
	(2.04)	(0.02)	(0.03)	(1.78)
Household wealth	-0.13	0.01	0.00	1.01
	(1.29)	(0.02)	(0.02)	(1.28)
Caregiver has never been to school	-1.61	0.03	0.02	-0.73
	(1.67)	(0.02)	(0.03)	(1.48)
Caregiver can read simple sentence	1.02	-0.02	-0.02	-0.63
	(2.52)	(0.04)	(0.05)	(2.63)
Books in house	0.79	-0.01	0.00	0.32
	(1.19)	(0.02)	(0.02)	(1.41)
Intervention (β_2)	47.26***	0.25***	0.18***	36.56***
	(1.67)	(0.03)	(0.02)	(1.75)
Comparison group mean	14.60	0.00	0.00	1.81
Number of observations	3,814	3,814	3,813	3,805

Table A.7 notes: this table shows results for estimating Equation 2 after adding the interaction terms shown here. This is an analog to Panel B of Table 6, adding the interaction terms shown here to test whether, for a reasonable set of observable controls, there is still a residual in the learning outcomes we study to be explained by aspirations which is not explained by the interaction of the intervention and other traits of the children and their families which also predict learning. We report clustered standard errors in parentheses below each estimated coefficient. $^{*}p < 0.10$, $^{**}p < 0.05$, $^{***}p < 0.01$.

Table A.8: How the mapping from baseline aspirations to endline learning changes with a large increase in the quality of educational supply, including interactions between intervention and various predictors of learning

	(1) Endline test score	(2) Child is literate	(3) Child is numerate	(4) Words read per minute
Educational aspirations x intervention (β_3)	1.02	0.06***	0.04*	3.28**
	(1.59)	(0.02)	(0.02)	(1.62)
Intervention x household wealth	-1.46	0.03	-0.01	-0.54
	(1.62)	(0.03)	(0.02)	(1.72)
Intervention x caregiver has never been to school	0.42	0.03	-0.01	1.57
	(1.48)	(0.03)	(0.03)	(1.74)
Intervention x caregiver can read simple sentence	-4.77**	0.06	-0.03	3.61
	(2.39)	(0.04)	(0.04)	(2.54)
Intervention x books in house	-2.21	-0.02	0.01	-0.56
	(1.79)	(0.03)	(0.03)	(2.21)
Educational aspirations (β_2)	3.39***	-0.00	-0.00	1.15**
	(0.94)	(0.00)	(0.00)	(0.50)
Household wealth	1.82*	0.00	-0.00	1.25*
	(1.02)	(0.00)	(0.00)	(0.64)
Caregiver has never been to school	-0.64	0.00	-0.00	-0.20
	(0.99)	(0.00)	(0.00)	(0.65)
Caregiver can read simple sentence	5.94***	-0.00	-0.00	1.38*
	(1.41)	(0.00)	(0.00)	(0.79)
Books in house	2.68***	0.00	-0.00	0.45
	(0.70)	(0.00)	(0.00)	(0.35)
Intervention (β_1)	49.05***	0.21***	0.18***	34.50***
	(2.49)	(0.05)	(0.05)	(2.99)
Comparison group mean	14.96	0.00	0.01	1.99
Number of observations	3,814	3,814	3,813	3,805

Table A.8 notes: this table shows results for estimating Equation 2 after adding the interaction terms shown here. This is an analog to Panel A of Table 6, adding the interaction terms shown here to test whether the intervention also has interactive effects with other observable household traits. We report clustered standard errors in parentheses below each estimated coefficient. $^{*}p < 0.10$, $^{**}p < 0.05$, $^{***}p < 0.01$.

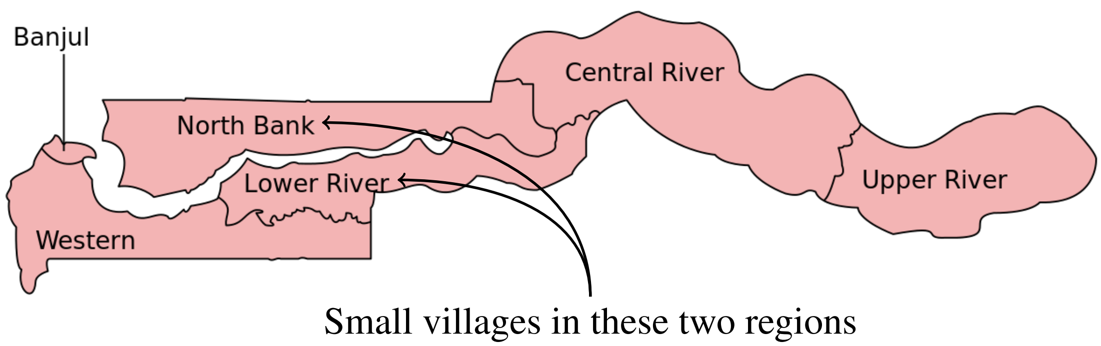
Table A.9: How the mapping from baseline aspirations to endline learning changes with a large increase in the quality of educational supply, including interactions between intervention and various predictors of learning

	(1) Endline test score	(2) Child is literate	(3) Child is numerate	(4) Words read per minute
Career aspirations x intervention (β_3)	-1.87	0.03	0.01	0.83
	(1.30)	(0.02)	(0.02)	(1.24)
Intervention x household wealth	-1.33	0.03	-0.01	-0.50
	(1.61)	(0.03)	(0.03)	(1.66)
Intervention x caregiver has never been to school	0.14	0.03	-0.02	1.32
	(1.49)	(0.03)	(0.03)	(1.73)
Intervention x caregiver can read simple sentence	-4.40*	0.07	-0.02	4.01
	(2.33)	(0.04)	(0.05)	(2.48)
Intervention x books in house	-2.49	-0.02	0.01	-0.72
	(1.76)	(0.03)	(0.03)	(2.17)
Career aspirations (β_1)	3.60***	0.00	0.00	1.27***
	(0.63)	(0.00)	(0.00)	(0.34)
Household wealth	1.70	0.00	-0.00	1.21*
	(1.04)	(0.00)	(0.00)	(0.66)
Caregiver has never been to school	-0.62	0.00	0.00	-0.19
	(1.00)	(0.00)	(0.00)	(0.64)
Caregiver can read simple sentence	6.02***	-0.00	-0.01	1.40*
	(1.41)	(0.00)	(0.01)	(0.76)
Books in house	2.89***	0.00	-0.00	0.52
	(0.69)	(0.00)	(0.00)	(0.34)
Intervention (β_2)	50.96***	0.23***	0.20***	36.08***
	(2.41)	(0.05)	(0.04)	(2.87)
Comparison group mean	14.60	0.00	0.00	1.81
Number of observations	3,814	3,814	3,813	3,805

Table A.9 notes: this table shows results for estimating Equation 2 after adding the interaction terms shown here. This is an analog to Panel B of Table 6, adding the interaction terms shown here to test whether the intervention also has interactive effects with other observable household traits. We report clustered standard errors in parentheses below each estimated coefficient. ${}^{*}p < 0.10$, ${}^{**}p < 0.05$, ${}^{***}p < 0.01$.

Figure A.1: Regions of The Gambia and study area

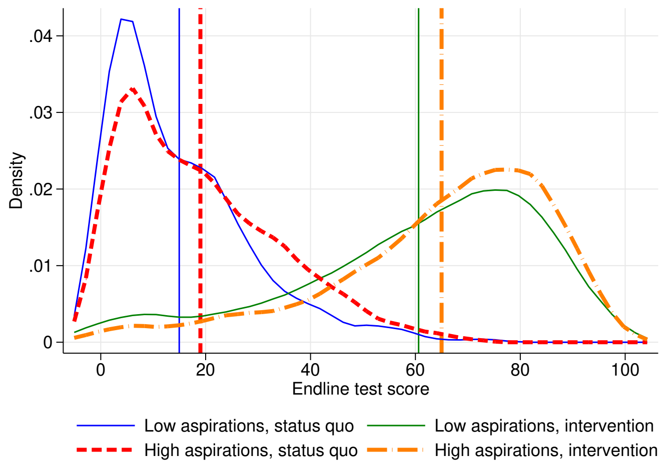
Panel A: The Gambia's location in West Africa



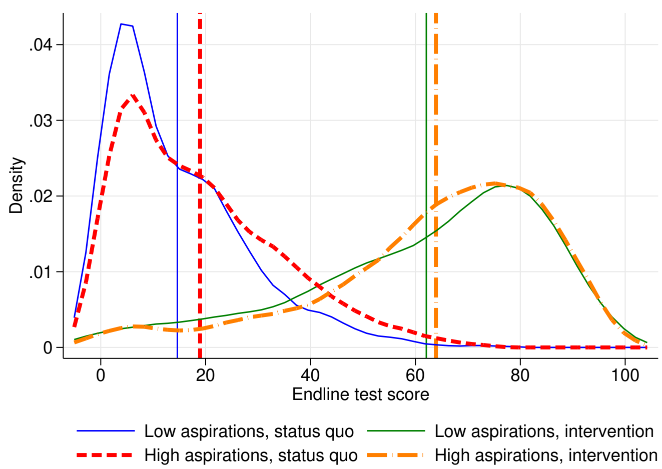
Panel B: Study area with The Gambia

Figure A.1 notes: this figure shows the location of our study area. In Panel A, we show a map of the continent of Africa with The Gambia shown within the red circle. In Panel B, we show a map of the Gambia, indicating the two regions where the study took place.

Figure A.2: Distributions of endline test scores, by baseline aspirations and status quo vs. intervention



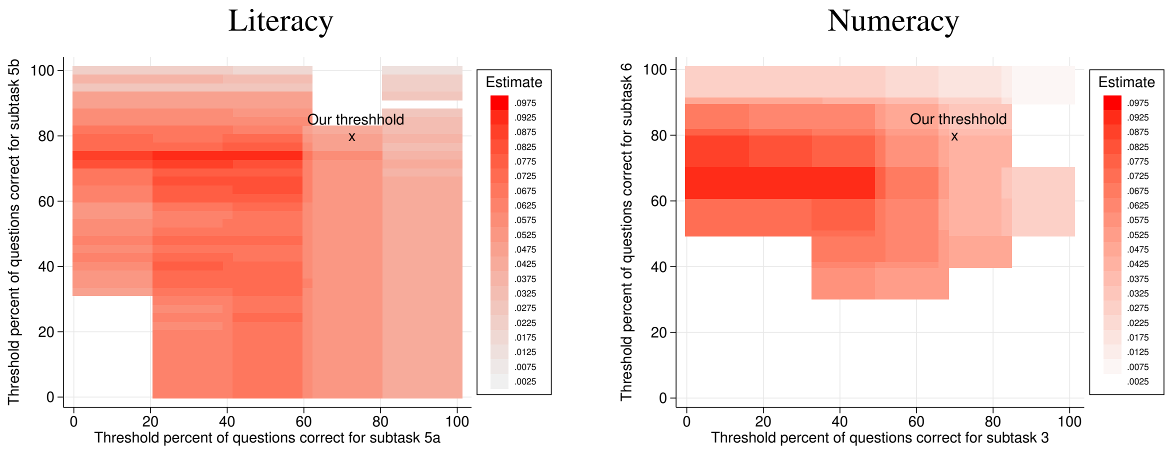
Panel A: Educational aspirations



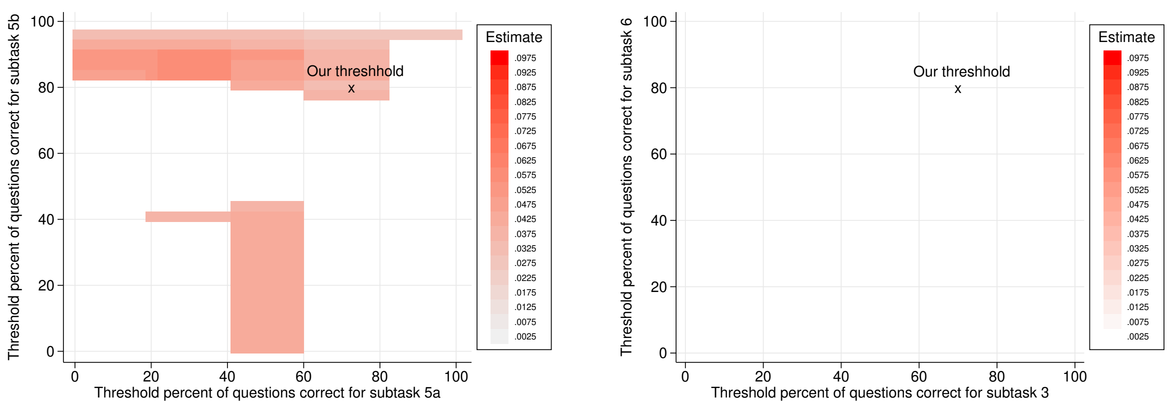
Panel B: Career aspirations

Figure A.2 notes: this figure shows kernel density plots of endline test scores for children whose caregivers did and did not express the aspiration listed in the panel title at baseline, and by whether or not they were enumerated in a village subsequently randomized to receive the intervention (that is, both the status quo and intervention), as indicated in the figure legends. The vertical lines show the mean test score of the group whose distribution is plotted with the same width, color, and pattern of line. All 3,813 observations in our estimation sample from Table 6 were used to generate these figures.

Figure A.3: Sensitivity analysis of β_3 across alternative definitions of literacy and numeracy, excluding estimates not significant at the 10 percent level



Panel A: Educational aspirations



Panel B: Career aspirations

Note: this figure shows heat maps of estimates of β_3 from Equation 2 for each skill (literacy or numeracy) and aspiration (education or career) combination. Each map plots the magnitude of estimates from each of 10,000 alternative definitions of literacy and numeracy. These 10,000 variables consist of each location on the 100-by-100 unit grid of all possible integer thresholds for the percent of questions answered correctly on each of the two subtasks comprising each skill (literacy and numeracy, respectively). We display as zero all results for which the estimate is not statistically significant at the 10 percent level. For reference, we plot the relevant threshold used (for either literacy or numeracy) in Table 6 with an x and overlay it on each graph.

For Online Publication: Appendix A: Test papers

Test papers begin on next page