When bootstraps aren't enough: Demand, supply, and learning in a very low-income context

Alex Eble and Maya Escueta*

February 2022

Abstract

Can families in low-income contexts "pull themselves up by their bootstraps?" In rural Gambia, caregivers with high aspirations for their children's future education and career, measured before the child starts school, invest substantially more in their children's education. Despite this, essentially none of these children are literate or numerate three years later. In contrast, a bundled supply-side intervention administered in these same areas generates large literacy and numeracy gains. Conditional on receipt of this intervention, children of high-aspirations caregivers are 25 percent more likely to achieve literacy and numeracy than those of low-aspirations caregivers. Our results show that even in very low-income contexts, greater caregiver aspirations for children can map onto substantially different child learning outcomes, but only in the presence of adequate complementary inputs.

*Eble: Teachers College, Columbia University, eble@tc.columbia.edu. Escueta: Sanford School of Public Policy, Duke University, maya.escueta@duke.edu. 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 Eble et al. (2021). This paper reports analysis of data collected in that study, 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, Naureen Karachiwalla, Josh Merfeld, Kate Orkin, Phillip Ross, and seminar audiences at the Columbia Committee on the Economics of Education, the Columbia Population Research Center, the Gui2de/World Bank joint seminar, the New York Nanodev conference, and the RISE 2021 Annual Conference. 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. Keywords: aspirations; education; poverty; complementarity.

1 Introduction

Many families wish to provide better lives for their children than experienced by previous generations. One central lever families use to achieve this goal is education. Intergenerational educational mobility has been shown to correspond to economic mobility (Black et al., 2011; Chetty et al., 2014, 2017), particularly in low and middle income countries (Azam and Bhatt, 2015; Asher et al., 2018; Alesina et al., 2021). A series of theoretical and empirical studies has shown a strong linkage from one specific type of desire for the future, known as "aspirations," to both greater investment in education and higher educational outcomes (cf. Beaman et al. 2012; Bernard et al. 2014; Genicot and Ray 2017; Lybbert and Wydick 2018; La Ferrara 2019).[1] It is not clear, however, whether this relationship – between desire for a better future for one's children, investment in their education, and greater educational outcomes – also holds in contexts where complementary inputs are absent, or of low quality.

In this paper, we study two core questions: first, if caregivers in low-income contexts want to raise their children's learning levels, how much learning can they bring about on their own? Second, how does this relationship change if we relax the constraint of very low-quality educational supply? To do so, we follow children and their caregivers in rural Gambia over a period of three years, beginning from the time immediately prior to the child starting primary school. We measure parent hopes for their child's future education and career, or "aspirations," at baseline. We use these as coarse measures of the family's desire to help the child towards a future that differs from what the family currently has. We estimate the mapping from baseline aspirations onto subsequent educational investment and learning. We contrast these relationships with the impact of a supply-side intervention which dramatically raises learning levels, and show how the aspirations-learning relationship in this context varies with a large increase in the quality of educational supply. We also show that the commonly-used test score standard deviation (SD) metric dramatically overstates the mapping from aspirations to learning when baseline levels of learning are extremely low.[2]

[1] See also the great summary of empirical work on this topic in Fruttero et al. (2021).

[2] The SD metric has been used for measuring learning gains in several hundred studies in the economics of edu-

We use data from a census of families in 169 villages in the two central regions of The Gambia. The data track families who, at the time of a baseline survey, intended to enroll at least one of their children in the first grade, for the first time, in the fall of 2015. This baseline survey collected families' educational and career aspirations for these children. Families and children were then followed over three years, during which time data was collected on the child's school enrollment, school-related time use, and on the family's educational expenditure for the child. At endline, these children were administered one-on-one tests of basic reading and math skills.[3]

We focus on caregivers' educational and career aspirations for the child, as in La Ferrara (2019). We operationalize this with two indicator variables, capturing the aspiration to have the child go to university, and the aspiration for the child to work in an urban area – a proxy for higher income jobs, given that income in the city is so much higher than that in the countryside. At baseline, sixty percent of families wish to send their children to university – roughly 10 percentage points lower than found in recent studies of aspirations in Ethiopia and Somalia (Bernard et al., 2014; Kipchumba et al., 2021) and almost 40 percent lower than the proportion of caregivers in rural India who aspire to send their child to junior college (Attanasio et al., 2020). A similar proportion aspire that their child will work in an urban area.

We find that higher baseline aspirations map onto greater subsequent investment in the child's education. Late enrollment in school is a common problem in Sub-Saharan Africa (Glewwe and Jacoby, 1995; Bommier and Lambert, 2000), and enrollment of children in school at younger ages is a strong predictor of greater overall educational attainment (Nonoyama-Tarumi et al., 2010). Caregivers with greater educational and career aspirations for their children are between three and six percentage points more likely to enroll these children in school in the first two years of the study, as compared to children of caregivers with lower aspirations. In the final year of the study, when there are no enrollment differences, high aspirations caregivers spend significantly more money on

cation and, in particular, reporting impact evaluations of educational interventions. For reference, see the numerous meta-analyses which collect and analyze these studies (Kremer and Holla, 2009; McEwan, 2014; Ganimian and Murnane, 2016; Glewwe and Muralidharan, 2016; Evans and Yuan, 2020).

[3]These were Early Grade Reading and Math Assessments, also known as "EGRA" and "EGMA" tests, respectively. See Platas et al. (2014) and Dubeck and Gove (2015) for details on their development, implementation, and limitations.

the child's education, and their children spend more time per day on school-related tasks.

Higher baseline aspirations also map onto higher endline test scores which appear, at first, to suggest substantially higher learning levels. Children whose caregivers report higher aspirations at baseline score 0.28-0.30 SDs better on a composite score of endline tests of basic reading and math ability, administered three years after the initial aspirations data were collected. The SD metric is a popular way to measure learning gains in studies of education, particularly in developing countries (Kremer and Holla, 2009; McEwan, 2014; Evans and Yuan, 2020). Put in the context of the hundreds of impact evaluations of educational interventions in developing country contexts covered by a series of recent meta-analyses[4], an estimate of 0.30 SD lies between the 75th and 90th percentile of effect sizes reported in these studies.

By comparing our effect size estimates to estimates of skill-based learning gains, we show that the SD metric dramatically overstates the relationship between baseline aspirations and subsequent learning. Specifically, using standard definitions of literacy and numeracy, we estimate a precise zero relationship between baseline aspirations and endline levels of either skill. We also find very small gains in other, lower-level skill attainment.

This contrast between the SD measure and absolute learning comes from the fact that, in our context, learning levels are compressed at the very left tail of possible values, despite our use of tests specifically designed to measure early skill acquisition, i.e., at this leftmost part of the skill distribution. From such a low baseline, even a very small absolute gain in test scores in this context translates into a large change in SD units; here, the 0.30 SD gain we measure translates to the child being somewhat more likely to master one additional rudimentary skill, such as differentiating which of two single-digit numbers is larger, or which of three words starts with a different sound. Such gains leaves essentially all children in this sample far from mastery of other basic skills necessary for literacy and numeracy – and expected of grade 2 students in The Gambia – such as reading simple words or calculating basic sums.[5]

[4]See those cited in the previous sentence and in Footnote 2.

[5]This inverse relationship between the learning contained in a given effect size estimate and the baseline learning level of the population being studied has also been found in US schools (Hill et al., 2008).

We argue further that our estimates provide a likely upper bound on the status-quo relationship between family motivation to improve children's lives, as measured by aspirations, and learning outcomes in this context. This is because potential unobservable confounders – for example, unobserved wealth or family preferences – are most likely to be positively correlated with both the aspirations we study and educational outcomes (Bernard et al., 2014; Ross, 2019). Should such unobserved traits influence our estimates, the true relationship would be even smaller than what we measure.

Next, we show that higher aspirations do map onto a much greater likelihood of achieving literacy and numeracy when high-quality complementary inputs on the supply side are present. As reported in Eble et al. (2021), a highly-resourced, bundled supply-side educational intervention randomly assigned to be offered in some of these same villages yielded transformative learning gains. In this paper, we show that this large increase in the quality of educational supply dramatically changes the relationship between baseline aspirations and endline learning.

Conditional on their village being randomly assigned to receive the intervention, children whose caregivers have high baseline educational aspirations for the child are 25 percent more likely to reach literacy and numeracy at endline than children in these same intervention villages whose caregivers do not. This finding echoes the results of a large RCT in Tanzania documenting evidence of complementarities between multiple supply-side educational inputs in increasing learning (Mbiti et al., 2019). We then analyze how aspirations interact with the intervention in the acquisition of individual skills at varying levels of difficulty. We find that the intervention is a substitute for caregiver aspirations in children's acquisition of low-level skills, and a complement to caregiver aspirations in children's acquisition of higher-level skills.

Finally, we discuss two potential alternative explanations for our estimates of how aspirations and the quality of educational supply interact to generate learning. These are that aspirations merely capture unobserved child ability or household wealth, respectively, which also lead to greater learning when the quality of educational supply increases. Unlike in our analysis of these relationships focusing on rural Gambian families not receiving the external intervention, we cannot

use a bounding argument to capture how unobserved factors may contribute to this second set of results. This is because the intervention could either substitute for or reinforce the role of these factors. Instead, we investigate the likely magnitude of these contributions. For child ability, a wide range of factors – including the extremely low proportion of caregivers who have ever gone to school or are able to read, the fact that aspirations are measured prior to the child starting school, and the fact that even after children go to school, caregivers in such contexts often have highly inaccurate beliefs about child ability (Dizon-Ross, 2019) – make it exceedingly unlikely that caregivers aspirations are merely a proxy for child ability. For wealth, we show that our main results are robust to including interactions between aspirations and measures of wealth and caregiver education, indicating that the interaction between aspirations and the intervention is not driven by this alternative explanation.

Our paper makes two key contributions. First, we advance understanding of how the demand-side and supply-side interact to generate learning in low-income contexts (cf. Jensen 2010; Glewwe and Muralidharan 2016; La Ferrara 2019; Muralidharan et al. 2019; Romero et al. 2020). We document that even in the face of severe income poverty, many families want to provide a better life for their children and act upon this desire by investing more in their children's education. We then show that these desires can map onto large gains in learning, but only when complementary inputs on the supply side are present. This adds to other recent evidence on the presence of complementarities in the provision of education in low-income settings (Mbiti et al., 2019; Kerwin and Thornton, 2021). Second, we show that in settings characterized by very low baseline learning levels, the test score standard deviation – a metric for measuring learning used in hundreds of previous studies (McEwan, 2014; Ganimian and Murnane, 2016; Evans and Yuan, 2020) – can yield an erroneously positive conclusion about the relative importance of educational inputs. In light of this, we argue that skill-based learning measures should be preferred when assessing the efficacy of interventions and the relative importance of various educational inputs in contexts with very low baseline levels of learning (Platas et al., 2014; Dubeck and Gove, 2015; Filmer et al., 2020). We 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; Serneels and Dercon 2021).

Our paper proceeds as follows. In Section 2 we describe our setting and data, and our measures of learning and aspirations. In Section 3 we present our reseach design. Section 4 shows our estimates of the mapping from aspirations to learning in the status quo of rural Gambia, how this varies by which measure of learning is used, and bounds these estimates given potential unobserved factors. Section 5 shows how this relationship changes with a substantial increase in the quality of educational supply and explores alternative explanations for these results. Section 6 concludes.

2 Background, setting, and data

In this section, we describe the setting in which we work, the data we analyze, and describe how we measure learning and aspirations.

2.1 Setting

Our study takes place in small, rural settlements in the North Bank and Lower River 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] It is a former British colony and served as a major hub for the trans-Atlantic slave trade. Its population is roughly two million people, and its geographic area covers roughly 11,300 square kilometers (CIA, 2019). It is also very income poor: per-capita GDP was estimated to be $716 in 2018. The country's main sources of economic activity are 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 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 even among children, learning levels are dramatically lower than in other countries in the region (Sprenger-Charolles, 2008).

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

Our study focuses on the population of children and families enrolled in the randomized controlled trial reported in Eble et al. (2021). The research took place in 169 villages in the two central regions of The Gambia, beginning with the universe of villages in these two regions who had between 10 and 300 households according to the 2013 national census.[7] All villages which met the predetermined eligibility criterion of having at least 10 eligible children were enrolled.[8] Children were eligible if, at the time of enumeration in early 2015, i) they were between the ages of 6 and 8, ii) they had not yet entered the first grade, and iii) their primary caregiver intended to enroll them in the first grade in the coming academic year. Because presence in this sample is conditional on the caregiver intending to enroll the child in school in the coming year, this means that aspirations measured among participants may differ from the population in these areas. When abstracting from our sample to the population of children in this age range in our study 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. For the sake of brevity, we focus on these 3,825 students in our analysis.[9] In the next section, we describe the characteristics of these children and their families.

2.2 Data

The data began with a census of children in study villages meeting the eligibility criteria described above. Data were collected from these children and their caregivers over the period from January 2015 to May 2018. Participants were enumerated in early 2015 and randomization occurred in late

[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.

Table 1: Demographic characteristics

	(1) All	(2) Status quo	(3) Intervention
Child is female	0.50	0.51	0.48
Child has five or more siblings	0.41	0.43	0.39
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		0.572	
(p-value)		(p= 0.684)	

Table 1 note: 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 by trial-assigned clusters of contiguous villages. All variables in this table, except for the number of observations, are binary, with 0 = No and 1 = Yes.

2015. Villages randomly assigned to the intervention arm received a highly-resourced, bundled intervention providing after-school remedial education delivered by para teachers. This program began in early 2016 and continued until the beginning of May 2018. In these villages, para teachers from within the village or nearby were hired and trained to use scripted lessons 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.[10]

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. From here onward,

[10]This intervention is described in greater depth in Eble et al. (2021).

we refer to children in villages randomized to not receive the intervention as the "status quo" group and those in villages randomized to receive the intervention as the "intervention" group. At baseline, more than three quarters of all children were being cared for by adults – whom we call caregivers – who had never been to school.[11] 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 are lower-income, more remote, and 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 Eble et al. (2021), with roughly one quarter of households living in homes constructed entirely out of synthetic materials.

We also collect three types of data on educational investment. The first captures child enrollment in school, collected at the end of each academic year. At the end of the third year, we also measure the child's school-related time use on an average weekday and, separately, caregiver expenditure on the child's education.

2.3 Measuring learning

The study measured learning via endline tests conducted in May and June of 2018. It used early grade reading and math assessments – EGRA and EGMA tests, respectively (Platas et al., 2014; Dubeck and Gove, 2015) – which were administered to all study children, one-on-one, as per the test guidelines. These tests are closely aligned with the Gambian national curriculum for grades 1-3, and versions of these tests have also been used as part of government teacher assessment efforts since 2007.

Each test is comprised of different "subtasks," i.e., skills. These skills capture different aspects of the process through which children acquire the skills to read with comprehension and calculate basic figures. In Table A.1 we describe the subtasks/skills evaluated by each test. There are six subtasks within each test; as the number of the subtask rises, so does the level of difficulty: for

[11] 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.

example, reading subtask 1 focuses on a skill, letter sound identification, believed to be much less difficult than that evaluated in reading subtask 4, familiar word recognition. We provide the full test papers in Appendix A.

We generate three key measures of learning using these tests. First, as in Eble et al. (2021), we generate a composite score of overall performance on the two tests. This is calculated as the average of the correct proportion of answers on each test, itself generated the average correct proportion of answers on each subtask. We estimate both the difference in raw scores between groups, as well as the transformation of this difference into units of the comparison group standard deviation of this score, henceforth the "SD" measure.

Our other two measures of learning draw on child performance on the individual subtasks within each test. First, we use standard, binary measures of literacy and numeracy based on performance on certain subtasks. A child is assumed to be literate if they can read "with good fluency" (45 words per minute) and correctly answer at least 80% of reading comprehension questions. A child is assumed 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, and correctly answer at least 80% of word problems. These definitions are the same used in Fazzio et al. (2021). Second, we show child performance on individual subtasks, as measured by the proportion of questions in that subtask answered correctly. A key advantage of the EGRA and EGMA tests is that they are designed to be able to measure even very early learning gains in the process of attaining literacy and numeracy. Measuring performance across these subtasks allows us to show detailed learning trajectories across a spectrum of skills, with differing levels of difficulty, on the path to these two benchmark abilities.

No baseline tests of learning were conducted. This decision was made at the end of pre-trial fieldwork in consultation with the Gambian Ministry of Basic and Secondary Education and other experts in the area. As we document, baseline learning levels are incredibly low in this area. Our fieldwork suggested that, because we are focusing 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 of participants – greatly exceeded the likely benefit of these tests, and we proceed assuming that everyone starts from a zero baseline learning level. The very low levels of learning at endline in the control group, after the vast majority of students have completed three years of primary schooling, further support this assumption.

2.4 Measuring aspirations

In the baseline survey, prior to randomization and also before the child would enter school for the first time, we collected caregiver aspirations for the child for the child's main caregiver. As in La Ferrara (2019), we focus on two types of aspiration. The first is the caregiver's aspirations for their child's highest level of educational attainment. The second is the caregiver's aspirations for their child's career in adulthood. To capture educational aspirations, we asked the child's main caregiver: "ideally, what is the highest level of education you would like [child name] to attain?" To capture career aspirations, we asked the caregiver: "when [child name] is 20 years old, what job do you hope [she/he] will be doing?"

These questions were designed to capture broader, latent variables about the caregiver's aspirations for the child's education and career, respectively. 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). Lybbert and Wydick's 2018 study of aspirations differentiates between "aspirational hope" and "wishful hope", arguing that the latter are characterized by a lack of a viable pathway to achieve them. In our study area in The Gambia, as in the Ethiopian, Indian, and Somalian contexts referenced above, few individuals are likely to go to college or university. Nonetheless, we follow these studies by referring to responses to the two questions as capturing aspirations. We see them as measures of a broader latent trait encompassing the caregiver's motivation to provide a future for the child different from what the family currently experiences. As we will show, they are also important predictors of the subsequent future-oriented investment behavior we study.[12]

[12]The aspirations we measure also differ importantly from expectations. In our pilot, we worked to choose language

In Table 2, we summarize caregivers' responses to these questions. In Panel A we show this for all children, and then for intervention children and status quo children, respectively, calculating the difference between the two groups. In Panel B, we conduct a similar analysis for girls and for boys. In Column 1, we first show the proportion of all children whose caregiver expresses the relevant aspiration for the child at baseline. We see that, at baseline, roughly 60 percent of caregivers report aspirations for the child to go to university. This is slightly lower than levels recently recorded in rural Ethiopia (Bernard et al., 2014) and Somalia (Kipchumba et al., 2021). The question regarding caregiver aspirations for their child's career allowed the respondent to answer freely; responses that did not clearly fall into one of 14 listed categories were recorded as given and later coded. Given economic conditions in The Gambia and our initial analysis of this response data, we considered the following two broad categories of employment: working for the government and working in an urban area. Since more than 85 percent of respondents chose some type of work for the government, leaving little variation to study, we instead focus on aspirations to work in an urban area.[13] This captures most jobs which require literacy and numeracy skills, and also takes in the fact that income in the city is much higher than in the countryside. Roughly 65 percent of caregivers express this aspiration for their child's career. The correlation between aspirations for the child to attend university and for the child to work in an urban area is 0.187, indicating substantial independent variation between the two.

Comparing across groups, we see no difference in baseline aspirations between intervention and status quo children. We also find small – no more than three percentage point – differences between caregivers' aspirations for girl and boy children for our main two aspirations variables. These differences are not statistically significant at conventional levels.

We next describe how baseline aspirations correlate with other baseline characteristics that might predict educational investment and learning levels. In Table 3, we present conditional means of aspirations levels by a series of variables related to relative economic prosperity, household

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

Table 2: Aspirations at baseline

Panel A: Control vs. treatment	(1) All	(2) Status quo	(3) Intervention	(4) P-value
Highest level of education: university	0.61	0.61	0.61	0.89
Broad work categories				
Job in urban area	0.65	0.65	0.65	0.93
Government job	0.87	0.86	0.87	0.70
Top jobs aspired to				
Teacher	0.26	0.26	0.27	0.89
Work for government (no further detail)	0.22	0.23	0.21	0.55
Nurse	0.15	0.15	0.15	0.94
Observations	3,825	2,045	1,780	—

Panel B: Female vs. Male	(1) All	(2) Female	(3) Male	(4) P-value
Highest level of education: university	0.61	0.60	0.63	0.06
Broad work categories				
Job in urban area	0.65	0.67	0.64	0.06
Government job	0.87	0.89	0.84	0.00
Top jobs aspired to				
Teacher	0.26	0.26	0.27	0.31
Work for government (no further detail)	0.22	0.20	0.24	0.00
Nurse	0.15	0.24	0.07	0.00
Observations	3,825	1,895	1,930	—

Table 2 note: this table presents baseline aspirations data for children in the trial, both overall (column 1) and, in columns 2 and 3, separately by the group as described in the panel title and column heading. Column 4 presents the p-value of a t-test for the null that the aspirations levels are equal at baseline for the groups mentioned in the panel title, derived from a regression of the variable on the panel group variable. We clustered standard errors by trial-assigned clusters of contiguous villages. These results are robust to adding controls for caregiver education and household wealth.

configuration, and caregiver education. We capture caregiver education with two variables: one is a binary variable that captures whether the child's caregiver has received any formal schooling; the other captures 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. We capture household wealth with a binary variable capturing whether the home is made of synthetic materials, as used in Eble et al. (2021) and Fazzio et al. (2021). We also include an indicator variable equal to one if the child has five or more siblings. This is meant to proxy for the number of children across which educational

investments must be divided. Finally, we show baseline levels of aspirations by whether there were any books found in the child's home during the baseline survey, and whether the child's main caregiver is their biological mother. These characteristics are all predetermined relative to our measurement of aspirations.

The child's caregiver is roughly ten percentage points more likely to express either of these aspirations for the child if the caregiver has previously been to school. We also see a large difference in baseline educational aspirations by whether the caregiver is literate or not (recall that only 8 percent of caregivers are literate), and a smaller difference by the presence of books in the house. There is some evidence of a difference in baseline career aspirations by household wealth, but not in educational aspirations. We see no differences in baseline aspirations by the number of siblings or the identity of the caregiver.

The pairwise correlations between aspirations and caregiver education are small (0.114 for the aspiration to have the child attend university, and 0.104 for them to work in an urban area). There is no evidence of correlation between our rough proxy of wealth and either aspirations measure, nor is there correlation between wealth and caregiver education. This is in line with the notion 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. In our later analysis of the relative importance of baseline aspirations on subsequent educational investment and learning gains, we add controls for these variables, isolating the relationship between the portion of aspirations orthogonal to these variables and our dependent variables.

3 Research design

In this section we describe the analyses we conduct to answer our core research questions. Our first research question asks how aspirations at baseline map onto later investment in education and endline learning levels for families in rural parts of The Gambia. To answer this research question, we estimate the following equation:

Table 3: Correlates of aspirations

	(1) Aspires that child will go to university	(2) Aspires that child will find work in urban area
Child gender		
Male	0.63	0.63
Female	0.60	0.68
P-value of difference	(0.08)	(0.04)
Caregiver education		
Caregiver has been to school	0.71	0.75
Caregiver has never been to school	0.58	0.63
P-value of difference	(0.00)	(0.00)
Caregiver literacy		
Cannot read simple sentence	0.60	0.64
Can read simple sentence	0.83	0.83
P-value of difference	(0.00)	(0.00)
Materials of home		
Home made of synthetic materials	0.62	0.68
Home made of natural materials	0.61	0.65
P-value of difference	(0.78)	(0.13)
Number of siblings		
Less than five	0.62	0.65
Five or more	0.63	0.67
P-value of difference	(0.61)	(0.42)
Number of siblings		
Less than five	0.57	0.65
Five or more	0.64	0.66
P-value of difference	(0.00)	(0.81)

Table 3 notes: This table shows the conditional means of aspirations across the baseline characteristics labeled in the left-most column. We break these baseline characteristics into binary variables, showing the conditional mean of the aspiration for both values of the characteristic, and the p-value for a t-test of the null that the aspiration in question is equal for those with each value of the baseline characteristic.

$$y_{itc} = \alpha_0 + \alpha_1 A_{t=0,ic} + \alpha_2 X_{t=0,ic} + \eta_r + \varepsilon_{ic} \quad (1)$$

This regresses the outcome variable of child i in cluster c at time t on α_0, a constant; $A_{t=0,ic}$, the aspirations of the caregiver for child i at baseline (i.e., when $t = 0$); $X_{t=0,ic}$, a vector of predetermined variables for child i, measured at baseline (these include all the variables shown in Table 3); and η_r, a region fixed effect. We cluster our standard errors at the level of contiguous clusters of villages, ε_{ic}.[14] Our main coefficient of interest is α_1, which captures the mapping from baseline aspirations to subsequent outcomes, after controlling for the baseline characteristics listed in Table 3, such as gender, wealth, and caregiver education.

In these analyses, we use only data from the status quo group. This is because, as shown in Eble et al. (2021), 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 context.

Our second research question asks whether the relationship from baseline aspirations to subsequent learning levels, estimated in question 1, changes with a dramatic improvement in the quality of educational supply. In these analyses we study children in both the status quo and intervention groups. Here we also use ordinary least squares, regressing the outcome variable on a constant, the treatment variable (which was randomly assigned), baseline aspirations, and their interaction:

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

Here T_c is child i's treatment status, and $A_{t=0,ic}$ is again the relevant measure of aspirations for the child reported by their caregiver at baseline. Here too we cluster our standard errors by cluster of contiguous villages. Our main coefficient of interest is β_3; the sign and significance of this coefficient indicates whether this change in the quality of educational supply changes the mapping from baseline aspirations to endline learning. Because the intervention is randomized and baseline

[14] This is the same level as the randomization in Eble et al. (2021).

aspirations are pre-determined, we do not include additional controls. Because our paper reports exploratory analysis of existing data, we did not register a pre-analysis plan (Olken, 2015; Lin and Green, 2016), though the analysis for the broader RCT was pre-specified and pre-registered (Boone et al., 2015).

4 Aspirations, educational investment, and learning

In this section, we show how baseline aspirations map onto subsequent educational investments and endline learning levels in the status quo group. The "investments" we consider are time and money. Time is measured by enrollment in school each year, as well as the proportion of time that the child spends on school-related tasks on a typical weekday in the final year of the study. The measure of monetary investment we use is the caregiver's educational expenditure on the child's education, also captured in the third and final year of the study. We study the mapping from baseline aspirations to endline learning levels first using the child's performance on the endline test. We then contrast this with estimates generated using standard measures of literacy and numeracy based on performance on these tests, as well as measures of the child's mastery of various specific reading and math skills. Finally, we bound our results, discussing the sign of potential influence from unobserved factors.

4.1 Aspirations and educational investment

We first characterize the mapping from baseline aspirations levels to subsequent educational investment. We present our estimates in Table 4; the outcome variables, named in the column headings, are educational expenditure in year three of the study, child time use in year three of the study, and enrollment in school in each of the three study years.

We find a positive and statistically significant mapping from both types of baseline aspirations to subsequent educational investments. For enrollment in school, we see evidence that children whose caregivers have higher educational or career aspirations for the child at baseline are 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: delayed enrollment in school is a strong predictor of lower overall educational attainment (Nonoyama-Tarumi et al., 2010).

For educational expenditure, we see that children whose caregivers hold higher educational or career aspirations for the child spend between 10 and 15 percent more per year on costs related to the child's education.[15] In addition, in that third and final year when essentially all children are enrolled in school, we find a statistically significant difference in the proportion of time on a typical weekday that the child spends on school-related tasks, with more time spent by children whose caregiver expressed high educational aspirations at baseline, though not for those expressing high career aspirations. The mapping from these aspirations to educational expenditure and enrollment are of the same order of magnitude as the estimated effect of the intervention-driven aspirations gain on educational investment measured in Bernard et al. (2014).

To better understand what other baseline characteristics predict early educational investment, we also present coefficients for other control variables. These show a few key patterns. First, the mappings from other factors to educational investment have a similar order of magnitude as do those for baseline aspirations. Second, the signs of these estimated relationships are as expected; for example, there is a statistically significant positive relationship between wealth and educational expenditure.

4.2 Aspirations and learning outcomes

We next characterize the relationship between baseline aspirations and endline learning levels, as measured by performance on the endline test, both in terms of the raw composite endline test score and standard measures of literacy and numeracy based on performance on given subtasks, as described in Section 2.2. We present these results in Table 5. In column 1 we show this for raw test scores. We find that, after controlling for baseline characteristics, children whose caregivers have higher baseline educational aspirations for the child perform 3.3 points better than children

[15] Expenditures are shown in 2020 US dollars; this difference is between 75 and 90 Gambian Dalasis, converted at a rate of 51.71 dalasis per dollar.

Table 4: Estimating the mapping from baseline aspirations to subsequent educational investment in the status quo group

	(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	79.04**	0.019**	0.032	0.052**	0.005
	(27.23)	(0.007)	(0.028)	(0.025)	(0.008)
Wealth index high	109.04**	0.003	-0.009	-0.029	-0.002
	(40.22)	(0.008)	(0.025)	(0.018)	(0.012)
Caregiver can read simple sentence	79.51	0.028**	0.066*	0.061**	0.017*
	(70.06)	(0.012)	(0.036)	(0.026)	(0.008)
Books found in house	72.49**	0.011*	0.047**	0.045***	0.007
	(28.23)	(0.007)	(0.020)	(0.014)	(0.006)
Child is female	-14.32	0.005	0.020	0.005	0.009
	(21.49)	(0.008)	(0.015)	(0.023)	(0.008)
Comparison group mean	611.36	0.545	0.825	0.802	0.971
Number of observations	1,862	1,908	1,937	1,909	1,908
Panel B: Career aspirations					
Aspiration: child will work in urban area	74.14**	0.005	0.036	0.055***	0.001
	(26.88)	(0.006)	(0.025)	(0.020)	(0.005)
Wealth index high	106.89**	0.003	-0.010	-0.030	-0.002
	(39.59)	(0.008)	(0.025)	(0.018)	(0.012)
Caregiver can read simple sentence	81.99	0.030**	0.067*	0.062**	0.017**
	(68.16)	(0.012)	(0.036)	(0.027)	(0.009)
Books found in house	77.20**	0.012*	0.049**	0.048***	0.007
	(28.00)	(0.007)	(0.019)	(0.014)	(0.006)
Child is female	-20.85	0.004	0.018	0.001	0.009
	(21.58)	(0.008)	(0.015)	(0.024)	(0.008)
Comparison group mean	617.54	0.553	0.820	0.799	0.973
Number of observations	1,862	1,908	1,937	1,909	1,908

Table 4 notes: this table shows the results of estimating Equation 1 on the outcome variables listed in the column heading using baseline aspirations as indicated in the panel heading. Dependent variables 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. $^{*}p < 0.10$, $^{**}p < 0.05$, $^{***}p < 0.01$.

Table 5: Estimating the mapping from baseline aspirations to endline learning in the status quo group

	(1) Endline test score	(2) Child is literate	(3) Child is numerate
Panel A: Educational aspirations			
Aspiration: child will go to college	3.278***	-0.001	-0.001
	(0.910)	(0.002)	(0.004)
Wealth index high	1.863*	0.001	-0.001
	(1.019)	(0.002)	(0.004)
Caregiver cannot read	5.985***	-0.001	-0.002
	(1.482)	(0.001)	(0.002)
Books found in house	2.960***	0.001	0.001
	(0.725)	(0.001)	(0.003)
Child is female	1.877**	-0.002	0.001
	(0.843)	(0.001)	(0.003)
Comparison group mean	14.964	0.001	0.006
Number of observations	1,971	1,971	1,970
Panel B: Career aspirations			
Aspiration: child will work in urban area	3.792***	0.002	0.003
	(0.658)	(0.001)	(0.004)
Wealth index high	1.771*	0.001	-0.001
	(1.034)	(0.002)	(0.004)
Caregiver cannot read	6.023***	-0.001	-0.002
	(1.468)	(0.001)	(0.002)
Books found in house	3.179***	0.001	0.001
	(0.711)	(0.001)	(0.003)
Child is female	1.583*	-0.002	0.001
	(0.878)	(0.001)	(0.003)
Comparison group mean	14.604	0.000	0.004
Number of observations	1,971	1,971	1,970

Table 5 notes: in this table we show results from estimating Equation 1 on the outcome variables listed in the column heading, for the baseline aspirations indicated in the panel title. These analyses include only children in the status quo group. The scale of the endline test score is 0-100. Literacy and numeracy are indicator variables. We report clustered standard errors in parentheses below each estimated coefficient. The full set of controls is as indicated in Section 3. $^{*}p < 0.10$, $^{**}p < 0.05$, $^{***}p < 0.01$.

of caregivers with low baseline aspirations group. Children in the low aspirations group answering 15 percent of questions on the endline test correctly (i.e., their score is 15 points). The difference for career aspirations is 3.8 points. These differences are both highly statistically significant.

We plot the distribution of these scores, by aspiration group, in Figure 1. This shows that the high aspirations group's test score distribution first-order stochastically dominates the that of the low aspirations group for both types of aspiration. Kolmogorov-Smirnov tests confirm this observation ($p<0.001$ in both cases). Using the common practice of scaling this raw difference by the standard deviation (SD) of the variable among the comparison group, this comprises a difference of 0.28-0.30.[16] These differences appear large in SD terms; they lie between the 75th and 90th percentile of effect estimates reported in the hundred of studies summarized in a series of recent meta-analyses of evaluations of educational interventions in such contexts, including Kremer and Holla (2009), McEwan (2014), Glewwe and Muralidharan (2016), and Evans and Yuan (2020).

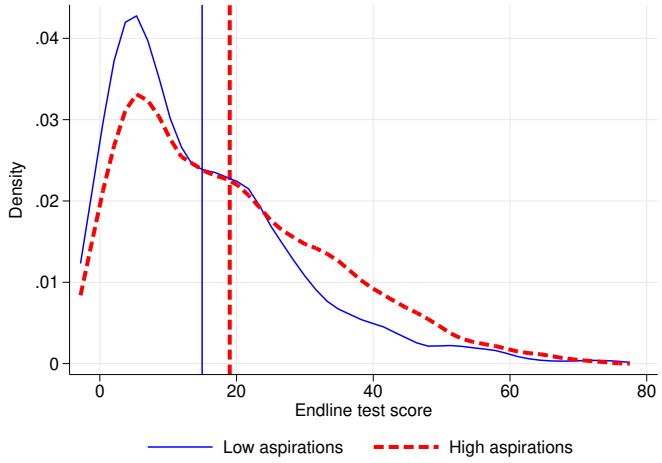
Next, we show that when measured in terms of skills gained, these gains appear much smaller and, for our measures of literacy and numeracy, disappear entirely. We take two approaches to measure skill acquisition. The first is to generate similar effect estimates, only using use binary measures of literacy and numeracy based on performance in the EGRA and EGMA tests described in Section 2.2. In columns 2 and 3, we report our estimates of Equation 1 using these two measures, respectively, as our outcome variables. We estimate precise zeros, finding no evidence of a relationship between baseline aspirations and endline likelihood of achieving literacy or numeracy. The confidence intervals we generate can reject anything larger than a one percentage point change in the likelihood of having attained either skill at endline as a result of having a caregiver with high educational or career aspirations at baseline.

The second way we measure the mapping of aspirations to subsequent skill acquisition uses test score gains in terms of the various skills each test evaluates.[17] In Figure 2, we plot the mean percent of correctly answered questions for skill by baseline aspirations levels. This illustrates

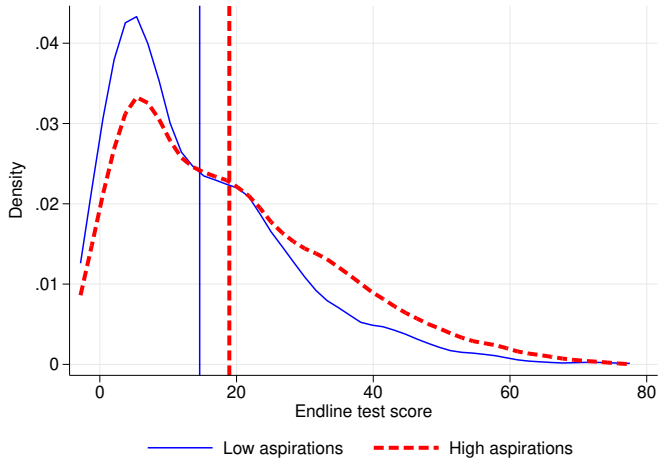
[16] Estimated using *Cohen's d*.

[17] See Table A.1 for a description of these skills and Appendix A for the full test papers.

Figure 1: Distribution of endline test scores for the status quo group, by aspirations



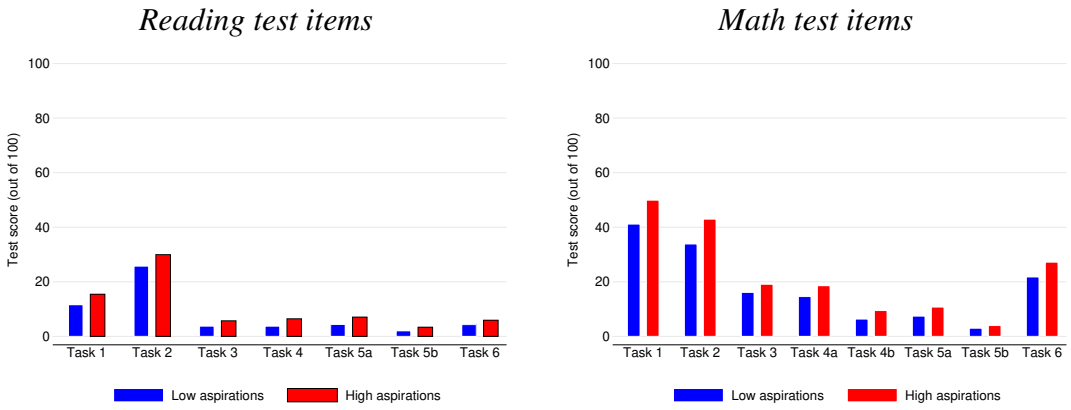
Panel A: Educational aspirations



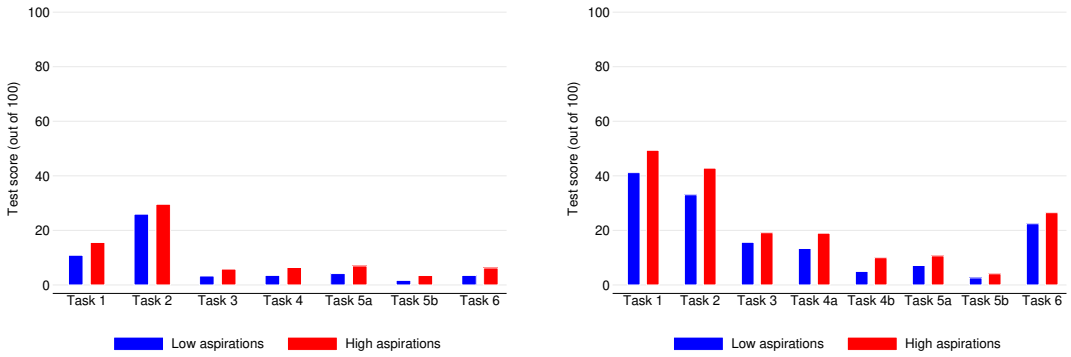
Panel B: Career aspirations

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. 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 $p \leq 0.001$.

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



Panel A: Educational aspirations



Panel B: Career aspirations

Figure 2 notes: This figure shows endline performance in the status quo group on each of the individual subtasks of the EGRA and EGMA tests, by baseline aspirations level. The relevant aspiration is indicated in the panel label. The number of observations per group are the same as in Table 5.

just how low endline skill levels are in the status quo group, regardless of baseline aspirations levels. It also highlights the very small absolute difference in skills between low aspirations and high aspirations children. In Tables A.2 and A.3, we show the regression equivalent of these comparisons, estimating Equation 1 using the subtask score as the dependent variable.

Our estimates of the high aspirations-low aspirations difference in endline test scores suggest that children whose caregivers hold high career or educational aspirations for the child would be more likely to master one more basic math or reading skill than children whose caregivers do not. In math, for example, this would mean these children would be able to differentiate which of two numbers was larger. In reading, it would mean that these children would be able to differentiate which of three words started with a different sound.

Unfortunately, all children in the status quo group display extremely low skill levels at endline. Figure 2 shows that, regardless of baseline aspirations, these children can correctly answer fewer than 10 percent of questions for most higher-level math and reading skills, such as single-digit subtraction or the ability to read simple, familiar words such as "and" and "but." As a result, even the large relative difference in test score standard deviations between children with low and high caregiver aspirations at baseline translates into a small absolute difference in endline reading and math skills.

This is best illustrated in the distribution of test scores, where we can see that essentially no children in the status quo group are even close to achieving either literacy or numeracy at endline. A helpful reference is that literacy and numeracy begin to manifest when a child correctly answers roughly 60 to 65 percent of questions on these tests. Examining the distributions in Figure 1, in both the high aspirations and low aspirations distribution of composite endline test scores, very few children score near these levels, with vanishingly few at or above them.

As a result, higher aspirations by themselves – even with the greater levels of investment that correspond to them – are unlikely to map onto a greater likelihood of literacy or numeracy for most children in rural Gambia. Furthermore, because in such contexts the school curriculum normally advances to higher level skills as children progress in grade level, even if these children do

not master the necessary lower-level skills (Pritchett, 2013; Muralidharan et al., 2019), the vast majority of these children are unlikely to ever achieve basic literacy or numeracy.

This finding builds on other work on how higher aspirations may not always lead to educational gains. Dalton et al. (2016) use a model to show that there can be suboptimally high aspirations, and 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 rural India. Leight et al. (2021) show that an intervention to raise aspirations in Ethiopia, similar to that studied in Bernard et al. (2014), has no measurable effect on either aspirations or investment. Here we have shown that even if higher aspirations do map onto greater investment in education, this does not correspond to meaningfully different learning trajectories in the status quo.

4.3 Measuring learning

Given the wide range of studies which use the SD metric to quantify learning gains (cf. McEwan, 2014; Ganimian and Murnane, 2016; Glewwe and Muralidharan, 2016), using effect sizes to compare the relative effectiveness of interventions or inputs across contexts is alluring and, in practice, common (Kremer and Holla, 2009; McEwan, 2014; Glewwe and Muralidharan, 2016; Evans and Yuan, 2020). Our findings in the previous section highlight another core result of our paper: in cases where learning levels are very low, using the test score SD metric can lead to incorrect conclusions about the importance of different learning inputs. This adds to prior work outlining the psychometric issues with the comparability of different tests (Furr, 2021).[18]

Our study emphasizes two core difficulties with using the SD metric to compare learning gains across contexts characterized by low baseline levels of learning. The first is that at very low levels of learning, a small absolute gain can be a large relative gain. For example, as we show in Figure 2, the 0.30 SD gain is comprised of many very small gains in skills. We estimate that the same increase in the percentage of questions answered correctly from a mean of 60 in the endline test

[18] Abhijeet Singh also describes these issues, as they relate to the economics of education in developing countries, in an excellent blog post here: https://blogs.worldbank.org/impactevaluations/how-standard-standard-deviation-cautionary-note-using-sds-compare-across-impact-evaluations, accessed November 1, 2021.

score would reduce the magnitude of this difference, measured in SD terms, by a third. Second, as we show later in the paper, the gain in skills such as literacy or numeracy acquired from a given increment in test scores varies dramatically with the child's position in the skill distribution. As we document, the 0.30 SD gain in learning we measure for the mapping from baseline aspirations to endline learning is larger than the vast majority of effect sizes measured in many hundreds of educational impact evaluations conducted in developing countries. At starting point where almost all children in our status quo group find themselves, however, it would require many multiples of that gain to generate a meaningful change in literacy.

This difference in interpretation is important. Were we to have relied only on test score standard deviations, we would have concluded that aspirations are a powerful predictor of learning gains. Using literacy, numeracy, and individual skill gains, however, we see that higher aspirations correspond to no greater likelihood of achieving literacy or numeracy, and only meager skill gains, during a crucial three-year period in these children's lives. In the context of US schools, Hill et al. (2008) report a similar pattern of greater effect sizes at lower grade levels (and thus lower levels of baseline skill), further highlighting the difficulty of using the SD measure for cross-context comparisons.[19]

There are several available tools for analyzing skill acquisition instead of test score SDs. The appropriateness of each tool will naturally depend on the context and research question. For assessment of low-level learning gains across contexts, the acquisition of early life skills such as familiar word reading or mastery of simple arithmetic tasks can be easily compared, particularly within the same language of assessment (Platas et al., 2014; Dubeck and Gove, 2015). For comparing across interventions or regions within a single national context, analysis of grade-specific skills, as per the expectations of the national curriculum, can be used (Muralidharan et al., 2019). For cross-context comparison of learning at higher levels of schooling, Filmer et al. (2020) propose a tool, learning-adjusted years of schooling (also known as LAYS), which allows for cross-context comparison of learning gains.

[19]Focusing on the US context, Kraft (2020) provides excellent guidance on how, and when, to use effect sizes for such comparisons.

4.4 Bounding our estimates for the status quo group

We argue that our estimates are likely to provide an upper bound for the true relationship between caregiver educational and career aspirations, educational investment, and child learning for these areas of rural Gambia. 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). 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 important, 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 larger in magnitude than the true relationship. This further emphasizes our main point that higher educational and career aspirations alone are likely insufficient to reach higher learning levels in this, and perhaps similar contexts.

As described in Section 2.2, presence in our sample is conditional on the caregiver intending to enroll the child in school in the coming year.[20] This means that aspirations may differ between the sample and the population of children in rural Gambia. Extrapolating to this latter group, we believe our estimates show a slightly different parameter. Specifically, our estimates of α_1 in Table 5 are an upper bound on what a child in this population might enjoy were their caregivers to possess high aspirations. Our argument rests upon the assumption that the children excluded from our study by the eligibility criterion of caregiver enrollment intent at baseline are likely to have either a similar or worse learning trajectory than those we included in the study. This premise is supported by multiple studies showing negative consequences in terms of learning and educational attainment stemming from late school enrollment in similar contexts (cf. Glewwe and Jacoby 1995; Bommier and Lambert 2000).

[20] 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.

5 Consequences of increasing the quality of educational supply

In this section, we show how the mapping from aspirations to learning outcomes changes when a key constraint – that of low quality educational supply – is relaxed. We also provide evidence on how demand, as measured by aspirations, and supply interact to generate learning at different levels of skill.

5.1 Aspirations, learning, and educational supply

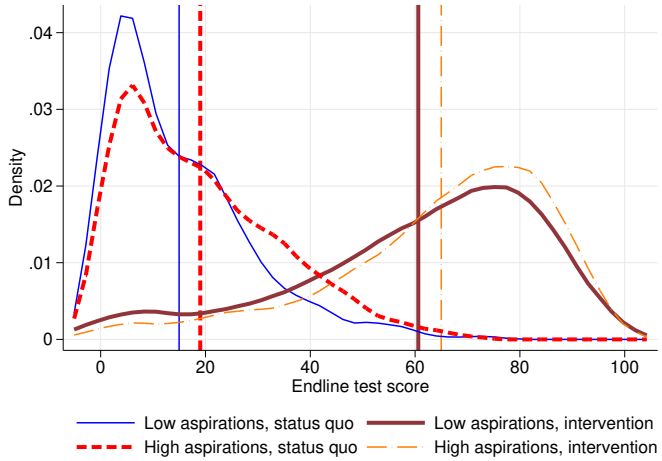
We first estimate how the mapping from baseline aspirations onto later learning outcomes changes when the quality of educational supply is dramatically improved by receipt of a bundled, supply-side intervention. In Figure 3, we plot the distribution of test scores among the four relevant groups – high and low aspirations children who did and did not receive the intervention, respectively. As in Figure 1, we show separate panels for educational and career aspirations. In both plots, and for both treatment assignments, we find that higher baseline aspirations map onto higher endline test scores.

We then estimate Equation 2 using data from the entire sample, i.e., both the status quo and intervention groups. We show these results in Table 6, using the three summary learning outcomes – standardized test scores, literacy, and numeracy – studied in Section 4. In Panel A, we show these results for educational aspirations; in Panel B, we show them for career aspirations.

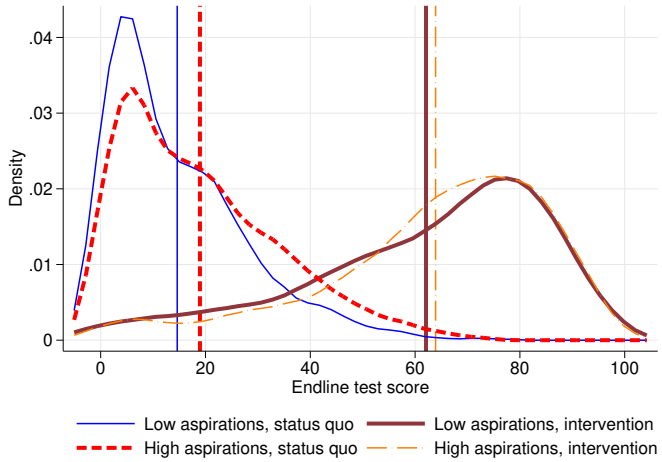
The first core finding is that the effect of the external intervention on learning dwarfs the mapping from higher baseline aspirations to learning. In terms of raw points on the test, the intervention generates a gain that is more than an order of magnitude larger than the low aspirations–high aspirations differential. In terms of literacy and numeracy, we estimate intervention-driven gains of of between 17 and 25 percentage points from a counterfactual case of essentially zero likelihood of demonstrating either skill.

The second core finding is that the mapping from baseline educational aspirations to endline literacy and numeracy is large and significant in the presence of high quality educational supply. As shown in Figure 1, in the rural Gambian status quo, the mapping from higher baseline aspirations

Figure 3: Distribution of endline test scores, by baseline aspirations and receipt of intervention



Panel A: Educational aspirations



Panel B: Career aspirations

Figure 3 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 within these groups, for children who were and were not resident at baseline in a village which was randomly assigned to receive the 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.

Table 6: How the mapping from baseline aspirations to learning varies in the presence of a large supply-side intervention

	(1) Endline test score	(2) Child is literate	(3) Child is numerate
Panel A: Educational aspirations			
Baseline aspirations x intervention	0.39	0.06***	0.04*
	(1.58)	(0.02)	(0.02)
Aspirations	3.65***	-0.00	-0.00
	(0.92)	(0.00)	(0.01)
Intervention	45.52***	0.23***	0.17***
	(1.74)	(0.02)	(0.02)
Comparison group mean	14.96	0.00	0.01
Number of observations	3,814	3,814	3,813
Panel B: Career aspirations			
Baseline aspirations x intervention	-2.44*	0.03	0.01
	(1.32)	(0.02)	(0.02)
Aspirations	3.86***	0.00	0.00
	(0.64)	(0.00)	(0.00)
Intervention	47.33***	0.25***	0.18***
	(1.68)	(0.03)	(0.02)
Comparison group mean	14.60	0.00	0.00
Number of observations	3,814	3,814	3,813

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 aspiration was 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 x 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. $^{*}p < 0.10$, $^{**}p < 0.05$, $^{***}p < 0.01$.

to greater endline test scores leaves the child very far from achieving literacy or numeracy. As we can see in both Figure 3 and Table 7, conditional on receiving the dramatic improvement in the quality of educational supply provided by the intervention, baseline educational aspirations lead to a four to six percentage point (or 24 to 26 percent) greater likelihood of attaining literacy and numeracy at endline. For career aspirations, however, we see a smaller, positive, but statistically insignificant relationship between baseline aspirations and literacy and numeracy, conditional on receiving the intervention.

5.2 Demand, supply, and skill acquisition

We next show how demand and supply interact to generate different levels of skill in this context. In Tables 7 and 8, we present estimates from using equation 2 to regress the acquisition of individual skills, as measured by the different subtasks in reading and math on each test, on aspirations, the intervention, and their interaction. Here again we focus on the β_3 coefficient – the interaction between baseline aspirations and receipt of the intervention. We also show a new parameter – the "interaction mean" – which is the sum of β_2 and β_3, and a p-value of a test of the null that it is equal to zero. The magnitude and statistical significance of this parameter tells whether, conditional on receipt of the intervention, the high-aspirations group demonstrates higher level of the skill in question. Our results show how the interaction of demand and supply in generating skill vary by skill difficulty. Specifically, our estimates of the interaction mean show that for the acquisition of simpler skills, the intervention appears to be a substitute for baseline aspirations in generating learning, while for more difficult skills, the two appear to be complements. In Panel A of both of these tables, we see a clear positive gradient: the sign and magnitude of our estimates effects of the interaction between baseline educational aspirations and the intervention change importantly with skill difficulty. For low-level skills, such as letter and number recognition (reading and math subtask 1, respectively), we estimate a statistically significant negative interaction term. For higher-level skills, such as familiar word reading and two-digit addition and subtraction, the estimates become positive and increase in magnitude. The confidence intervals around the lower-

Table 7: Demand, supply, and reading skill acquisition

	Subtask 1	Subtask 2	Subtask 3	Subtask 4	Subtask 5a	Subtask 5b	Subtask 6
Panel A: Educational aspirations							
Baseline aspirations x intervention (β_3)	-1.745	-1.044	2.322	3.418	3.191	4.264**	-0.651
	(1.936)	(2.057)	(1.885)	(2.223)	(2.122)	(2.075)	(2.273)
Baseline aspirations (β_2)	3.349**	4.095***	1.648*	2.421***	2.400***	1.226**	1.052
	(1.276)	(1.272)	(0.841)	(0.841)	(0.866)	(0.543)	(0.872)
Intervention (β_1)	55.737***	24.628***	45.475***	57.575***	54.428***	42.227***	56.889***
	(2.143)	(2.082)	(1.874)	(2.227)	(2.251)	(2.075)	(2.436)
Interaction mean ($\beta_2 + \beta_3$)	1.604	3.051	3.970	5.839	5.591	5.490	0.401
P-value [$\beta_2 + \beta_3 = 0$]	[0.274]	[0.067]	[0.022]	[0.005]	[0.005]	[0.008]	[0.851]
Comparison group mean	37.820	37.261	25.238	30.705	29.915	21.682	31.135
Number of observations	3,683	3,683	3,683	3,683	3,683	3,683	3,683
Panel B: Career aspirations							
Baseline aspirations x intervention (β_3)	-4.123**	-0.633	-1.305	-1.228	-1.176	0.503	-1.368
	(1.566)	(1.940)	(1.641)	(1.736)	(1.621)	(1.804)	(2.296)
Baseline aspirations (β_2)	4.355***	3.504***	2.332***	2.482***	2.386***	1.664***	2.422***
	(0.878)	(1.252)	(0.562)	(0.597)	(0.582)	(0.380)	(0.685)
Intervention (β_1)	57.332***	24.367***	47.727***	60.433***	57.114***	44.489***	57.382***
	(2.011)	(2.097)	(1.963)	(2.071)	(2.054)	(2.076)	(2.414)
Interaction mean ($\beta_2 + \beta_3$)	0.232	2.871	1.027	1.254	1.210	2.167	1.054
P-value [$\beta_2 + \beta_3 = 0$]	[0.857]	[0.043]	[0.504]	[0.441]	[0.424]	[0.220]	[0.625]
Comparison group mean	37.656	37.404	25.632	31.591	30.752	22.279	30.365
Number of observations	3,683	3,683	3,683	3,683	3,683	3,683	3,683

Table 7 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 A.1. We report clustered standard errors in parentheses below each estimated coefficient. The tests are shown in their entirety in Appendix A, divided by subtasks and with subtask 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 8: 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								
Baseline aspirations x intervention (β_3)	-5.304**	-4.032*	3.385*	0.660	2.873	1.619	7.364***	-0.193
	(2.273)	(2.242)	(1.729)	(1.955)	(2.194)	(1.880)	(2.474)	(2.031)
Baseline aspirations (β_2)	7.462***	7.650***	2.276**	3.431***	2.859***	3.070***	1.046	5.416***
	(1.983)	(1.827)	(0.894)	(1.070)	(0.788)	(0.811)	(0.681)	(1.131)
Intervention (β_1)	49.649***	49.716***	41.106***	46.540***	56.763***	39.007***	47.004***	26.923***
	(2.866)	(2.786)	(1.773)	(2.128)	(2.169)	(1.611)	(2.260)	(1.972)
Interaction mean ($\beta_2 + \beta_3$)	2.158	3.618	5.661	4.091	5.732	4.689	8.410	5.223
P-value [$\beta_2 + \beta_3 = 0$]	[0.049]	[0.006]	[0.000]	[0.015]	[0.006]	[0.007]	[0.001]	[0.002]
Comparison group mean	64.822	57.450	35.478	36.491	32.851	25.710	24.955	34.343
Number of observations	3,682	3,682	3,682	3,682	3,682	3,682	3,682	3,682
Panel B: Career aspirations								
Baseline aspirations x intervention (β_3)	-6.480***	-7.478***	-1.931	-3.558*	-2.731	-2.445	2.114	-2.499
	(2.026)	(1.909)	(1.560)	(2.002)	(2.184)	(1.700)	(2.418)	(2.000)
Baseline aspirations (β_2)	7.279***	8.805***	3.029***	5.531***	4.826***	3.500***	1.527***	4.058***
	(1.713)	(1.576)	(0.749)	(1.045)	(0.847)	(0.682)	(0.533)	(1.040)
Intervention (β_1)	50.567***	52.063***	44.406***	49.237***	60.274***	41.558***	50.108***	28.376***
	(2.694)	(2.544)	(1.877)	(2.145)	(2.405)	(1.725)	(2.517)	(2.081)
Interaction mean ($\beta_2 + \beta_3$)	0.799	1.327	1.098	1.973	2.095	1.055	3.641	1.559
P-value [$\beta_2 + \beta_3 = 0$]	[0.454]	[0.208]	[0.421]	[0.246]	[0.296]	[0.496]	[0.125]	[0.360]
Comparison group mean	65.094	57.615	36.279	36.287	33.057	26.468	25.887	35.698
Number of observations	3,682	3,682	3,682	3,682	3,682	3,682	3,682	3,682

Table 8 notes: this table shows results for estimating Equation 2 for children's scores on the individual math 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 A.1. We report clustered standard errors in parentheses below each estimated coefficient. The tests are shown in their entirety in Appendix A, divided by subtasks and with subtask 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. $^{*}p < 0.10$, $^{**}p < 0.05$, $^{***}p < 0.01$.

level subtasks (1 and 2 in both reading and math) exclude the point estimates for the higher-level subtasks (4 and 5), and vice versa. For the two most difficult subtasks – reading and math subtasks 5b; reading comprehension and two digit subtraction with borrowing, respectively – we estimate a large and statistically significant positive interaction term.[21] The interaction mean itself is small and insignificant for the lower level subtasks, and grows in magnitude and statistical significance as we approach the most difficult subtasks.

We can see this even more clearly by adding the coefficients on aspirations and the aspirations x intervention, i.e., $\beta_2 + \beta_3$ from Equation 2. Using the estimates from column 1 in Panel A of either Table 7 or 8, we see that adding the two generates a much smaller and statistically insignificant parameter. This indicates that conditional on receipt of the intervention, children of high-aspirations caregivers are no more likely to perform better on this subtask than children of low-aspirations caregivers. Looking at the estimates from the most difficult subtask – subtask 5b in both tests – we see that the interaction is statistically significant, large, and positive. This indicates that at higher levels of skill, aspirations and the intervention complement each other.

This reveals an important feature of the way that aspirations map onto learning. In both the status quo and intervention group, high-aspirations families are trying to push out on the frontier of learning. Figure 2 and Tables A.2 and A.3 show that, in the status quo group, the largest differences in endline learning between the children of low- and high-aspirations caregivers occur at the lowest level skills: subtasks 1 and 2 in both reading and math. The intervention greatly advances that frontier, so that the largest learning differences we see in the status quo group are erased in the intervention group. The intervention greatly raises the platform from which high-aspirations families are trying to reach, allowing differences in learning levels to appear between children of low- and high-aspirations caregivers in the intervention group in the most difficult subtasks, where before (Figure 2 and Tables A.2 and A.3) there were none. In Panel B of both tables, we see evidence of the first pattern - that career aspirations and the intervention are also substitutes in the production of lower-level skills – but do not appear at higher level skills. This

[21] 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.

is consistent with the notion that career aspirations differ from educational aspirations in terms of how they are acted upon, as we saw in column 2 of Table 4

Unlike in Section 4.4, it is much more difficult to bound our estimates in this section using the likely sign of effects from other contributing sources. This is because the intervention could either substitute for inputs from other sources that might be correlated with our measure of aspirations – such as household wealth or unobserved child ability – or complement them. If these inputs were complements, our estimates would be an upper bound, as the isolated mapping from aspirations and the intervention would likely be smaller. If they were substitutes, our estimates would likely be lower bounds.

Instead, we argue that the difference between the estimates here and those in Section 4 in terms of order of magnitude and significance reveal the relative importance, and complementary nature, of supply and demand in this context. They show three key messages. First, they show that our very coarse measure of demand, by itself, is correlated with investment behavior, and the acquisition of basic skills, but ultimately very little high-level learning. Second, they show that adequate supply can tremendously increase learning independent of our measure of demand. Third, they show that in the presence of adequate supply, our measure of demand is correlated with a substantially greater likelihood of children reaching both literacy and numeracy, due to the intervention moving out the frontier from which high-aspirations families invest and push their children. As a whole, they show that even in a very low-income context, demand can matter, but only in the presence of adequate supply.

5.3 Alternative explanations

We next address two potential alternative explanations for our key results, focusing on potential contributors to our measures of aspirations. The first is correlation between unobserved child ability and aspirations. The second is correlation between unobserved family wealth and aspirations.

There are several reasons why the correlation between unobserved child ability and aspirations is highly unlikely to be the main explanation for our results. First, in this context and, particularly,

at the time that we measure aspirations, caregivers are highly unlikely to know whether the child is of high ability. At the time these data are collected, the child had not yet been to school. Furthermore, more than three quarters of caregivers in these data have never been to school, and over 90 percent of them cannot read a short, simple sentence. As a result, it is highly unlikely that they are able to identify academic skill among their children at the time when baseline aspirations are measured, prior to the child's first ever enrollment in school. As Dizon-Ross (2019) documents, even after enrollment in school, caregivers in low-income contexts often have highly inaccurate beliefs about child ability. Second, the educational investment behavior of caregivers corresponds to aspirations in a way that is consistent with the investment response to an experimentally-generated increase in aspirations as measured in Bernard et al. (2014). Third of all, while career and education aspirations both predict subsequent investment behavior, they are only mildly correlated (correlation 0.18). Finally, we can look within families at how much aspirations vary, as a proxy for unobservable (to us) within-family differences in ability that may manifest as aspirations differences. 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, in the vast majority of cases, our aspirations data capture family desires for their children's futures, rather than serving as a proxy for family beliefs about individual child ability.

Similarly, there are several reasons why it is highly unlikely that some broader, latent socioeconomic variable is what drives the interaction between baseline aspirations and the supply-side intervention. First, we see evidence of baseline educational aspirations leading to greater likelihood of literacy and numeracy in the presence of the intervention, but no such relationship for career aspirations. Second, we conduct a robustness test which estimates an alternative version of Table 6 after adding interactions between the intervention and household wealth, caregiver education, caregiver literacy, and the presence of books in the home. We present these results for baseline educational and career aspirations in Tables A.4 and A.5, respectively. These show that the main

patterns we observe in Table 6 are robust to the inclusion of these other predictors of a potential response to the presence of the intervention. In other words, for a reasonable set of observable controls, we show that there is still a residual in the learning outcomes we study to be explained by educational aspirations which is not explained by the interaction of the intervention and these other traits of the children and their families which also predict learning. As in Tables 7 and 8, our estimates for career aspirations show no evidence of positive interaction effects, underscoring the difference between educational and career aspirations.

6 Conclusion

In this paper we study a common desire that many families across the world have: the aspiration that their children will go on to live better lives than those lived by previous generations. We characterize the relationship between this desire, subsequent educational investment, and later learning in a very low-income context. We first show that many families in this context posses this desire and act on it, investing more in their children's education even in the face of the extremely high opportunity cost of investment. We show that these investments can lead to a much greater likelihood of achieving literacy and numeracy by the end of the third grade, but only when complementary resources on the supply side are present. This finding belies the notion that families in such contexts merely need to wish and try to "pull themselves up by their bootstraps."

We also show that, when baseline learning levels are very low, the widely used measure of test score standard deviations can lead to incorrect conclusions about how educational inputs do or do not generate meaningful differences in learning. We find that, in this context, the test score standard deviation generated a spuriously optimistic assessment of the relationship between aspirations and learning outcomes. We argue that in such contexts with very low counterfactual learning levels, skill-based measures of learning gains may be more accurate in capturing the importance of various educational inputs and the impact of interventions.

Overall, our research highlights an important feature of the educational experience of children and their families in extremely resource-poor contexts such as the one we study. As is the case

across the world, many caregivers in our sample wish to improve the life chances of their children and help them to reach a prosperous adulthood. We show that these caregivers expend dear financial resources to do so, both in terms of money and their children's time. These investments yield a statistically significant return in terms of the child's performance on literacy and numeracy tests, relative to that of their peers. Sadly, because counterfactual learning levels are so low, these relative gains still leave children dramatically short of reaching literacy and numeracy, 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, these same aspirations map onto far greater likelihood of the child being able to read with understanding and conduct basic arithmetic. For policy, this suggests that while the demand side can yield important learning gains in low-income contexts, 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.

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.

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.

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.

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.

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.

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?," *Center for Global Development, Working Paper*, 2020, *545*.

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).

Ferrara, Eliana La, "Aspirations, Social Norms, and Development," *Journal of the European Economic Association*, 2019, *17* (6), 1687–1722.

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.

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.

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.

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.

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.

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

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.

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: Exper-

imental evidence from Tanzania," *The 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.

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.

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.

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.

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.

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

Appendix

Table A.1: Test subtasks

Reading		*Math*	
Subtask	**Example**	**Subtask**	**Example**
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 __)
		4a	Simple addition (e.g., 3+2)
4	Read a familiar word (e.g., but)	4b	Two- and three-digit addition (e.g., 38+26)
5a	Read a short passage	5a	Simple subtraction (e.g., 5-3)
5b	Answer questions on the passage's content	5b	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 A.1 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.

Table A.2: Mapping of aspirations at baseline to individual reading subtask performance 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	3.132** (1.311)	3.664*** (1.275)	1.516* (0.854)	2.380*** (0.872)	2.210** (0.894)	1.267** (0.559)	1.123 (0.890)
Comparison group mean	11.592	25.741	3.744	3.729	4.371	2.028	4.309
Number of observations	1,971	1,971	1,971	1,971	1,971	1,971	1,971
Panel B: Career aspirations							
High baseline career aspirations	4.177*** (0.850)	3.172** (1.253)	2.292*** (0.559)	2.620*** (0.598)	2.383*** (0.576)	1.794*** (0.347)	2.343*** (0.655)
Comparison group mean	10.884	25.949	3.295	3.499	4.183	1.671	3.494
Number of observations	1,971	1,971	1,971	1,971	1,971	1,971	1,971

Table A.2 notes: this table shows results for estimating Equation 1 for children's scores on the individual reading subtasks. 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 A.1. The tests are shown in their entirety in Appendix A, divided by subtasks and with subtask 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 individual math subtask performance 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	6.804*** (2.001)	7.031*** (1.842)	2.104** (0.922)	3.072*** (1.095)	2.652*** (0.776)	2.770*** (0.844)	0.847 (0.645)	5.112*** (1.143)
Comparison group mean	41.153	33.866	16.109	14.594	6.337	7.414	2.978	21.779
Number of observations	1,970	1,970	1,970	1,970	1,970	1,970	1,970	1,970
Panel B: Career aspirations								
High baseline career aspirations	6.670*** (1.713)	8.180*** (1.564)	2.941*** (0.742)	5.215*** (1.026)	4.660*** (0.793)	3.230*** (0.652)	1.428*** (0.504)	3.857*** (1.037)
Comparison group mean	41.183	33.074	15.623	13.371	4.958	7.132	2.597	22.450
Number of observations	1,970	1,970	1,970	1,970	1,970	1,970	1,970	1,970

Table A.3 notes: this table shows results for estimating Equation 1 for children's scores on the individual math subtasks. 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 A.1. The tests are shown in their entirety in Appendix A, divided by subtasks and with subtask 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: How the mapping from baseline educational aspirations to learning varies in the presence of a large supply-side intervention, adding interactions with various other predictors of learning

	(1) Endline test score	(2) Child is literate	(3) Child is numerate
Baseline educational aspirations x intervention	0.32	0.06***	0.04*
	(1.53)	(0.02)	(0.02)
Baseline educational aspirations x household wealth	2.30	0.01	0.02
	(1.71)	(0.03)	(0.03)
Baseline educational aspirations x caregiver has never been to school	-0.00	-0.05*	-0.04**
	(1.43)	(0.03)	(0.02)
Baseline educational aspirations x caregiver can read simple sentence	-1.85	0.01	-0.05
	(2.89)	(0.05)	(0.05)
Baseline educational aspirations x books in house	3.71***	0.02	0.05***
	(1.26)	(0.02)	(0.02)
Baseline educational aspirations	0.75	0.02	-0.00
	(1.66)	(0.03)	(0.02)
Household wealth	-0.36	0.01	-0.02
	(1.31)	(0.01)	(0.02)
Caregiver has never been to school	-0.27	0.05**	0.02
	(1.19)	(0.02)	(0.01)
Caregiver can read simple sentence	5.16**	0.02	0.02
	(2.42)	(0.04)	(0.04)
Books in house	-0.50	-0.02	-0.02
	(1.23)	(0.02)	(0.02)
Intervention	45.58***	0.23***	0.17***
	(1.71)	(0.02)	(0.02)
Comparison group mean	14.96	0.00	0.01
Number of observations	3,814	3,814	3,813

Table A.4 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.5: How the mapping from baseline career aspirations to learning varies in the presence of a large supply-side intervention, adding interactions with various other predictors of learning

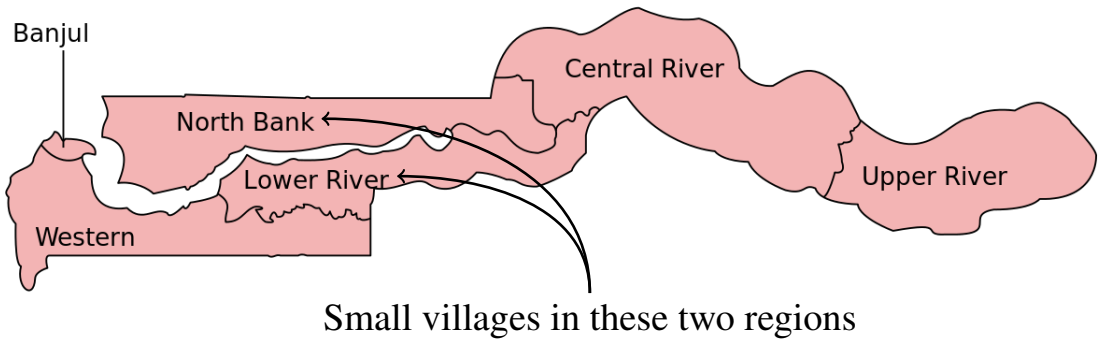
	(1) Endline test score	(2) Child is literate	(3) Child is numerate
Baseline career aspirations x intervention	-2.39*	0.03	0.01
	(1.32)	(0.02)	(0.02)
Baseline career aspirations x household wealth	1.65	0.00	-0.02
	(1.48)	(0.02)	(0.02)
Baseline career aspirations x caregiver has never been to school	1.77	-0.03	-0.04
	(1.95)	(0.02)	(0.03)
Baseline career aspirations x caregiver can read simple sentence	3.93	0.06	0.01
	(3.13)	(0.04)	(0.05)
Baseline aspirations x books in house	1.58	0.00	0.00
	(1.40)	(0.02)	(0.02)
Baseline career aspirations	0.77	0.02	0.04
	(2.04)	(0.02)	(0.03)
Household wealth	-0.13	0.01	0.00
	(1.29)	(0.02)	(0.02)
Caregiver has never been to school	-1.61	0.03	0.02
	(1.67)	(0.02)	(0.03)
Caregiver can read simple sentence	1.02	-0.02	-0.02
	(2.52)	(0.04)	(0.05)
Books in house	0.79	-0.01	0.00
	(1.19)	(0.02)	(0.02)
Intervention	47.26***	0.25***	0.18***
	(1.67)	(0.03)	(0.02)
Comparison group mean	14.60	0.00	0.00
Number of observations	3,814	3,814	3,813

Table A.5 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. $^{*}p < 0.10$, $^{**}p < 0.05$, $^{***}p < 0.01$.

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.

For Online Publication: Appendix A: Test papers

Test papers begin on next page