174

Working Paper

Young Lives

Starting Together, Growing Apart:

Gender Gaps in Learning From Preschool to Adulthood in Four Developing Countries

Abhijeet Singh and Sofya Krutikova

www.younglives.org.uk

NOVEMBER 2017

Starting Together, Growing Apart: Gender Gaps in Learning From Preschool to Adulthood in Four Developing Countries

Abhijeet Singh and Sofya Krutikova

ISBN 978-1-909403-97-0

A catalogue record for this publication is available from the British Library.

Core-funded by

UKaid from the British people

Young Lives, Oxford Department of International Development (ODID), University of Oxford, Queen Elizabeth House, 3 Mansfield Road, Oxford OX1 3TB, UK

Tel: +44 (0)1865 281751 • Email: younglives@younglives.org.uk

Contents

The authors

Abhijeet Singh is an Assistant Professor at the Stockholm School of Economics and a Research Associate of the Young Lives study, with which he has been associated since 2008. His research applies micro-econometric methods to the study of salient issues of policy interest in developing countries, focusing especially on issues relating to the economics of education and child health and the delivery of public services.

Sonya Krutikova works at the Institute of Fiscal Studies as Programme Director of the Centre for the Evaluation of Development Policies (EDePo). Her main research interests are in the determinants of skill acquisition among children and young people living in poverty, as well as more broadly the mechanisms through which childhood conditions manifest in child development and outcomes. Her recent work focuses on the role of home and school factors in explaining the evolution of gaps in cognitive skills and school attainment among children from poorer and better off backgrounds in developing countries.

Acknowledgements

We thank Young Lives for providing the data in this paper and supporting this study. We are grateful to Abhijit Banerjee, Sonia Bhalotra, Raquel Fernandez, Deon Filmer, Chris Karbownik, Diane Schanzenbach and seminar participants at the World Bank, the University of Oxford (CSAE), LSE (APPAM conference) and FAU Nuremberg for their comments on earlier drafts of this paper. We thank Maria Franco Gavonel and Jalnidh Kaur for research assistance and Juan Leon for assistance in the generation of comparable test score measures.

Summary

This working paper studies the evolution of gender gaps in multiple cognitive skills from the ages of 5 to 19 years old, using Young Lives unique panel data from Ethiopia, India, Peru and Vietnam; it is the most extensive panel-based investigation on this question in developing countries. The findings suggest that, in all four countries, gender gaps in learning are either absent or small in absolute magnitude prior to school entry (at 5 years old) and at primary school age (8 years old). Larger gaps emerge later, widening particularly between the ages of 12 and 15; gaps favour boys in Ethiopia, India and Peru, but girls in Vietnam. This is in contrast to OECD contexts, where significant gender gaps in maths and language skills tend to be in the same direction. Subsequently, these learning gaps appear to mostly persist until early adulthood. In establishing the direction, magnitude, and persistence of gender gaps, we pay careful attention to issues of ordinality and decay in test scores. Panel-based, value-added models with a rich set of covariates including past achievement, child health, time use, parental education and wealth, and school quality, explain at most half to two-thirds of the cross-sectional gender gap in test scores at 15 years old.

1. Introduction

Gender-based gaps in education have long been of concern to economists and policymakers. This is for several reasons: inequalities in human capital may directly translate into later inequalities in labour force participation, the nature of employment and wages; if arising from factors unrelated to productivity, such inequalities could indicate a misallocation of resources; and, perhaps most importantly, equality of opportunity, regardless of gender or other social markers, remains a valuable policy objective in its own right.[1]

In developing countries, the core focus on gender-based inequalities in education has typically related to enrolment and grade progression through school. In these areas, considerable progress has been made in the past 15 years.[2] However, years of schooling may hide substantial systematic differences in the actual levels of skill development in children.[3] Thus valid, and increasingly important, questions remain about the presence, extent and sources of gender gaps in test scores within and across countries.[4]

Although fundamental to understanding whether and where these gaps are most of concern, and to identify the ages and dimensions in which interventions might most be required, current evidence on these questions is particularly scarce in developing countries. This largely reflects the limited availability of suitable data. Internationally comparable household-based surveys, such as the Demographic and Health Surveys or the Living Standards Measurement Studies, collect details on the enrolment and current grade of individual children in the household, but do not administer tests of learning. Comparative school-based international assessments, such as PISA and TIMSS, have limited coverage of developing countries and, further, exclude students who are not enrolled in schooling or absent on the day of the assessment. This selection is particularly a concern in contexts where both enrolment and attendance may vary systematically by gender. They are also limited in the range of background information that they collect. These concerns are also true of other school-based assessments, including national test scores. Finally, in individual developing countries, the paucity of panel datasets with measures of achievement has led to these sources being substantially understudied compared to, for example, in the US.[5]

1 For an early statement of similar grounds of concern, see Mill (1869, Chapter 4).

2 For example, the 2015 UNESCO Global Monitoring Report documents that 'gender disparity in primary enrolment has been substantially reduced since 1999, but not eliminated' (UNESCO 2015: 155). It further states: 'Countries where gender gaps have been reversed underline the dynamic nature of achieving gender parity. Careful analysis of these trends is needed to inform future policy' (166).

3 This is evident from a long literature which looks at the effects on achievement of different schools in the same setting (e.g. charter schools or Catholic schools in the US, or private schools in developing countries) or, in the cross-country case, on the differential human capital implied by a year of schooling across different countries (e.g. Schoellman 2012, 2016; Singh 2017).

4 For instance, a recent paper highlighting the sharp reduction in gender inequality in access to schooling notes: 'Schooling attainment, as measured by grades of school completed, does not necessarily accurately reflect the learning outcomes of children, particularly in contexts of social promotion to the next grade level and large variations in school quality and in family background ... These differences may have implications for gender differences in learning despite the same level of schooling attainment if girls are likely to attend different types of schools than boys, tend to take different classes than boys, are treated differently than boys in the same classes, or are treated differently outside of school than boys are.' (Grant and Behrman 2010: 87)

5 Most developing countries do not have the high-quality administrative datasets that would allow for consideration of these questions even in the absence of external panels (see the analyses by Figlio et al. (2016) using Floridan administrative data). We are not aware of any such analyses in a non-OECD country.

This working paper addresses this gap using Young Lives unique panel data from Ethiopia, Andhra Pradesh state in India, Peru and Vietnam, where Young Lives has followed two cohorts of children from 2002 to 2014. The paper focuses on three key questions. First, is there a gender gap in achievement and how, if at all, does it differ in magnitude and direction across countries and different domains of learning such as quantitative and language abilities? Second, how do these gaps evolve in each domain over the course of childhood – at what ages do they first emerge, and do they then substantially decline or increase with age? Third, what are the proximate sources of these gaps – can their emergence perhaps be explained by observed differences in household investments, child endowments, time use in different activities, or the quality of school attended?

The data used, collected by Young Lives, present several particular strengths for this analysis. Foremost, they cover a long age range from 5 to 19 years old (preschool age to early adulthood), with comparable tests of achievement across countries and ages in multiple learning domains. Since the data are collected through home visits of a birth cohort, they do not suffer from selection due to school enrolment or attendance on the day of testing. The panel dimension of the data allows for the analysis of learning dynamics. In particular, we can account for the extent to which gender gaps observed at any particular age may be accounted for by differences in achievement that were already evident at earlier ages. Further, we can estimate value-added models of achievement which, compared to cross-sectional specifications, allow for a more robust investigation of various sources of gender-based divergence in achievement. Finally, the four countries represent very different cultural contexts, with important differences in gender-related attitudes and social norms, and thus are likely to provide a broad spectrum of gender-based differences in test scores in developing countries.

We document three main descriptive patterns in our analysis. First, we do not find much evidence of large gender gaps in learning at school entry age (5 years old) or in early primary school (8 years old) in any country. Gender gaps do, however, develop at later ages in most countries and are particularly evident after the age of 12, a period coinciding with adolescence and post-primary schooling. Second, there is important heterogeneity in the direction of gender gaps, where significant gender gaps mostly favour boys in Ethiopia, India and Peru, but typically favour girls in Vietnam. There is also important heterogeneity in magnitude: while gaps are typically modest and frequently insignificant in Peru, they are larger in other countries and most striking in India and Vietnam. Third, in contrast to most developed countries, we typically find lower evidence of heterogeneity across domains of learning in our study countries: where significant, gender gaps in mathematics and vocabulary are consistently in the same direction.

Finally, we document significant differences in household investment, enrolment and other factors determining learning between boys and girls which, along with prior test scores and the quality of schools enrolled in, help to partially account for the emergence of gender gaps. The extent to which we can explain the emergence of the gaps with these observable characteristics and test scores differs across countries/tests and, typically, we can explain between half and two-thirds of the cross-sectional gender gap at 15 years old.

Our results relate to a large literature studying gender-based inequalities in both developed and developing countries, including in academic achievement, to which we contribute in multiple dimensions.[6] Most importantly, we are the first to be able to study gender gaps in test scores over such a long age range in multiple developing countries. Comparing results across ages highlights that whereas educational outcomes for boys and girls show few signs of systematic bias at primary school ages, gender gaps are frequently prominent after the age of 12 and, having emerged, often persist until adulthood. This suggests that between 12-15 years old, a period marking the important transitions into post-primary education and adolescence, may be particularly crucial for gender-based divergence in achievement and, as such, for interventions seeking to moderate eventual gaps in adulthood.

Second, we pay greater attention to methodological concerns relating to the measurement of student achievement, intergroup comparisons, and dynamic analyses than most previous work. In particular, we take seriously the challenges arising from the ordinality of test scores and decay in student achievement in the analysis and interpretation of test score gaps over time. Recent evidence, mostly from the US, indicates that these issues are of first order importance for assessing the direction, magnitude and persistence of intergroup differences. However, these have not previously been investigated in the context of learning inequalities in developing countries.

Third, in analysing potential sources of the gap, we are able to account for a much broader range of possible channels than previous studies. This is feasible only because the data have substantial information on past achievement, child health, time use, parental education and wealth, and the sorting of different boys and girls across schools. This compares especially favourably with the limited background information in cross-sectional international assessments, such as PISA and TIMSS, upon which previous work has been based.

Finally, the substantial heterogeneity in the direction and magnitude of gender gaps in learning across contexts and ages, which we highlight, is directly relevant for policy discussions regarding inequalities in learning. In particular, these discussions are often not sensitive to such variation and implicitly assume that gaps favour boys and are of substantive magnitude.[7] This presumption could be misleading for education policy priorities where the gaps are in the opposite direction (as in Vietnam). While understanding such heterogeneity is clearly of great importance for establishing policy priorities, the paucity of suitable internationally comparable data sources in developing countries means even basic facts in this area are not well-established.

The rest of the paper is structured as follows: Section 2 describes the data used; Section 3 presents descriptive evidence on the magnitude of cross-sectional gender gaps in achievement and then examines these in a panel setting to shed further light on the ages at which these gaps emerge; Section 4 focuses on the age window of 12-15 years old, which appears to be critical for the emergence of gender gaps, and attempts to explain statistically

---

6 See Niederle and Vesterlund (2010) and Fryer and Levitt (2010) for a discussion of previous work on gender gaps in test scores. The closest studies to our paper are Fryer and Levitt (2010) and Bharadwaj et al. (2012). These focus solely on the gender gap in mathematics, with panel-based analyses restricted to the US and Chile, respectively, and for shorter periods of childhood. Fryer and Levitt (2010) present panel-based analyses only for students from kindergarten to Grade 5 in the US, Bharadwaj et al. (2012) study gaps between Grades 4 and 8 in Chile which, although an OECD member and classified as a high-income country by the World Bank, is perhaps closer to the contexts we consider. Both papers do present international comparisons of gender gaps, but are necessarily restricted to cross-sectional analyses using 15 year olds enrolled in schooling in the countries covered by international assessments.

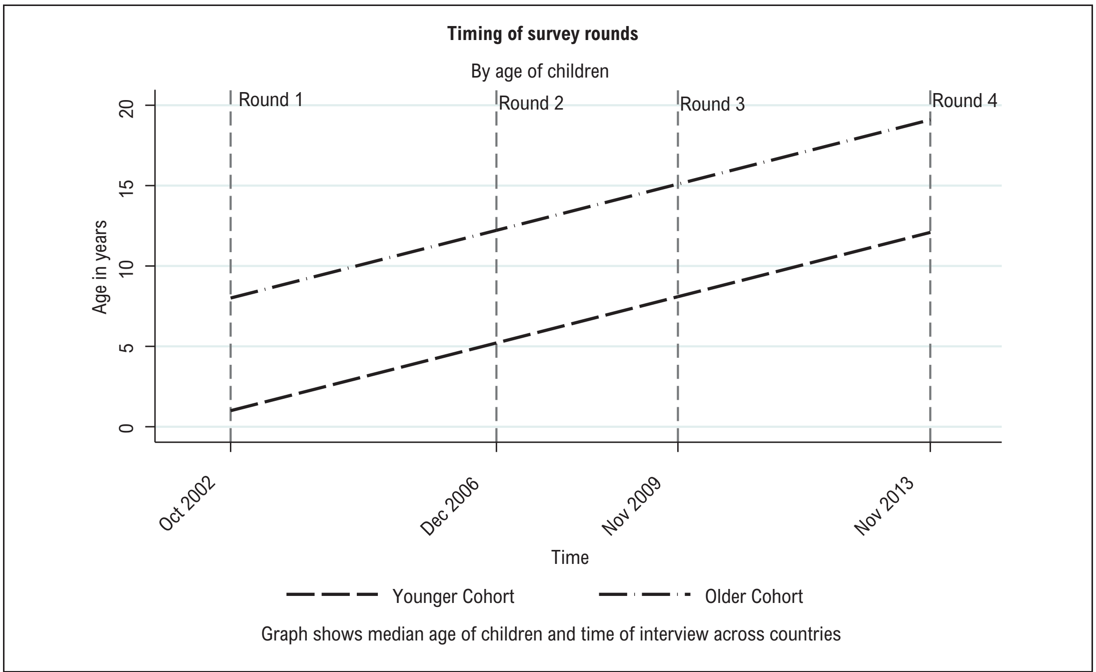
7 For an example of such policy focus, see the United Nations Girls Education Initiative, the related £340 million Girls Education Challenge programme funded by UKAid through the Department for International Development, or the Nike Foundation's The Girl Effect programme.

the emergence of these gaps to differences in home environments, time allocation and the schooling of boys and girls; Section 5 concludes.

2. Data

Data used in this paper come from the Young Lives longitudinal study, which has followed two cohorts of children in four countries – Ethiopia, Andhra Pradesh state in India, Peru and Vietnam – over four waves between 2002 to 2013/14.[8] The Older Cohort, of about 1,000 children each in the four countries, was born in 1994/95 while the Younger Cohort (about 2,000 children in each country) was born in 2001/02. Figure 1 shows the children's ages at the time of each survey round. In this paper, we use data from Rounds 2-4, observing the Younger Cohort children at 5, 8 and 12 years old, and the Older Cohort at 12, 15 and 19 years old.[9] The survey tracks children who migrated in later rounds from their initial community in the 2002 round and attrition rates in the sample are very low, with over 90 per cent of the sample still in the 2013/14 round.[10]

Figure 1. *Age of Young Lives sample individuals in successive survey rounds*



8 The study is ongoing, with a further round of data collection in 2016-17, the data from which are not yet released in the public domain. During this study, the state of Andhra Pradesh (with a population of 84 million people in 2011) was bifurcated into Telangana and Andhra Pradesh states in 2014. Throughout this paper, when referring to Andhra Pradesh, we mean the undivided state as it existed until 2014. In terms of enrolment and learning outcomes, Andhra Pradesh is typically close to all India averages (see Pratham 2015). In the paper we refer to results for 'India' or 'the Indian sample'; readers should keep in mind that the sample is exclusively based in this one state.

9 The first round of the survey administered only minimal assessments of learning to children in the Older Cohort, then aged 8. The Younger Cohort, then aged ~12 months, were not administered any cognitive assessments.

10 Attrition from all causes excluding deaths between the 2002 and 2013 rounds is under 5 per cent for the Younger Cohort in all countries except Peru, where it is 6.3 per cent. Attrition in the Older Cohort from all causes excluding deaths ranges from 4.3 per cent in Andhra Pradesh to 11.3 per cent in Vietnam. A detailed breakdown of attrition is available at www.younglives.org.uk.

2.1 Household data

In each survey round, detailed questionnaires were administered regarding various household characteristics and child-specific information. These include standard demographic and socio-economic information such as household structure, parental education, access to services and wealth but also, importantly for the purpose of this paper, extensive information on the individual child including time use, expenditures on the child's education, and their nutritional status (measured using WHO anthropometric scores).

2.2 Tests administered in Young Lives

Young Lives has administered a wide variety of tests in various rounds, summarised in Box 1. Quantitative skills are assessed at 5 years old using the orally administered quantitative subscale of the Cognitive Development Assessment (CDA) for preschool-aged children. At later ages (8, 12, 15 and 19 years), quantitative skills are assessed using paper-based mathematics tests. Given wide variations in the grade and skill levels of individuals both within and across countries, the tests are not designed to be grade appropriate and incorporate questions at widely differing levels of difficulty. A substantial subset of items in the quantitative assessments is common across countries and age groups from 8-19 years old.

Language skills are measured using a battery of different tests over the study period. Receptive vocabulary is tested using adapted versions of the Peabody Picture Vocabulary Test in the four countries between the ages of 5 and 15 years old. At the age of 15, in Round 3 of the survey (2009), language skills were additionally tested using a cloze test where students complete a sentence by providing a missing word ('fill in the blanks'). Reading comprehension was measured in Round 4 of the survey (2013/14) using a language-specific reading comprehension test delivered to 12 and 19 year olds.

2.3 Item Response Theory test scores

Box 1. *Cognitive tests in Young Lives*

Cohort	Round 2 (2006)	Round 3 (2009)	Round 4 (2013)
Younger Cohort	5 years old Receptive vocabulary CDA Quantitative	8 years old Receptive vocabulary Mathematics	12 years old Receptive vocabulary Mathematics Reading
Older Cohort	12 years old Receptive vocabulary Mathematics	15 years old Receptive vocabulary Mathematics Cloze test	19 years old Mathematics Reading

Notes: CDA refers to the Cognitive Development Assessment quantitative subscale.

Test scores used in this paper are constructed using Item Response Theory (IRT) models. These models posit a relationship between a unidimensional latent proficiency parameter and the probability of answering a question correctly. It is assumed that the relationship is specific to the item, but is constant across individuals. The use of IRT models is standard in international assessments such as PISA and TIMSS, albeit less so in economics of education research in developing countries.[11]

11 For a detailed explanation of IRT models and their estimation, see Van der Linden and Hambleton (1997) and Das and Zajonc (2010).

There are three key advantages for using IRT scores in this paper. Most importantly, using common items administered across age/round/country samples as 'anchor items' for the linking of tests scores, it is possible to put scores for different samples on the same metric. This enables an assessment of whether the absolute magnitude of gender gaps increases and decreases with age or time, or whether it is larger or smaller in one context compared to another. Linking in IRT models assumes that item characteristics are population invariant.[12] While reasonable for mathematics tests across countries/rounds/cohorts, this is untenable for vocabulary and reading tests across languages. Hence we link maths scores across countries from 8-19 years, but only link vocabulary and reading scores within language.[13]

Second, by allowing test questions to differ in their characteristics such as difficulty, IRT scores provide a less arbitrary measure of achievement than commonly reported percentage correct scores. This is important because it is possible that gender gaps in achievement are particularly concentrated on questions of particular difficulty levels: aggregation of test scores which provide an equal weight to all questions may then provide a misleading estimate of gender gaps.[14]

Finally, IRT models allow for a better range of diagnostics to assess the comparability of assessments across contexts, ages or time. This is important because students in particular settings may be more familiar with certain types of questions or modes of testing. While not a concern for comparing gender gaps within a sample, this is central to being able to compare across ages, time or countries.

IRT models only identify ability up to a linear transformation and therefore require normalisation. In this paper, we normalise scores to have a mean of zero and a standard deviation of one in the base age group in which it is administered. Specifically, mathematics scores are normalised with reference to the distribution of test scores of 8 year olds pooled across countries; the receptive vocabulary scores are normalised with reference to the 5-year-old age group within language; and the reading scores, available only for 12 and 19 year olds, are normalised with reference to the 12-year-old age group within language.[15]

2.4 Ordinality of test scores

A key issue, common to all test scores, is that they are ordinal measures and not measured on an interval scale; any monotonic transformation of test scores is conceptually an equally valid test score (see Bond and Lang 2013, 2017; Nielsen et al. 2015). This is not a problem solved either by the use of IRT models or by applying common procedures for standardisation. Ordinality of the outcome measure greatly complicates the study of

12 Concretely, this imposes that the relationship between latent proficiency and the probability of answering a question correctly is the same for all individuals in the population. For example, children of identical ability in Vietnam and in India should have the same probability of getting a correct answer on a given question.

13 This non-comparability of reading and vocabulary scores across languages is not a problem for within-country analyses of gender gaps; it merely entails that magnitudes of gaps should not be compared across countries.

14 As an example of how this may matter in practice, Singh (2015) documents that the causal private school effect on English scores in rural India looks considerably larger when using IRT scores, in the metric of standard deviations of the score, than a standardised raw score. Dividing questions by the task required, Singh further demonstrates that the private school effect is larger on 'harder' questions than on 'easier' ones, thus making issues of weighting particularly salient.

15 Scores were generated using the OpenIRT software package written by Tristan Zajonc on the pooled datasets. Performance of the model in explaining variation was assessed by inspecting the fit of each individual test item to the Item Characteristic Curve. Differential Item Performance was similarly judged; where indications of DIF were found, for example across countries, we split the item in the estimation of the IRT models.

intergroup differences. It renders results on the magnitude of gaps suspect and, if the Cumulative Distribution Functions (CDFs) of test scores for two groups cross, then even the direction of the gap can be reversed by arbitrary rank-preserving scaling decisions.

We take these issues seriously and attempt to address them upfront to the extent possible. First, rather than restricting our cross-sectional comparisons within sample to just documenting differences in the mean of test scores, we look at the entire distributions of scores. If distributions do not differ across groups, we can be reassured that any finding of no gaps in mean achievement does not reflect scaling decisions. Additionally, for all cross-sectional comparisons, we present the CDFs of the full distribution of test scores for boys and girls: where there is a gender gap, but the distribution for one group first order stochastically dominates the other, we can be reassured that the direction of a gap is invariant to rankpreserving transformations of the test metric. This approach relies only on the ordinal crosssectional information in test scores, as recommended by Bond and Lang (2013).

Second, in a panel setting when we are looking at the divergence of test scores across ages, we will in all cases investigate non-parametrically the differences in test score trajectories between boys and girls. Specifically, for any two successive age points, we will non-parametrically predict scores in the later period based on initial scores. If these trajectories do not cross – that is, the predicted score for Group A exceeds the predicted score for Group B across the full distribution of prior achievement – this indicates that our conclusions about the age periods in which divergence occurs are invariant to rank-preserving transformations of the baseline scores.

Third, we always concentrate on levels of test scores as an outcome, which may be more easily treated as approximately on an interval scale, rather than changes in test scores where such an interpretation is less justified.[16] This is most clearly evident in our descriptive analyses above, but is also clear in our regression-based analyses attempting to account for gender gaps – specifically, all our regressions will be specified with levels of current and past achievement on the left and right hand side respectively, rather than changes in the test scores (i.e. adopting 'dynamic OLS' value-added models rather than 'gain score' specifications, in the terminology of Guarino et al. (2015)). To our knowledge, this is the most careful consideration of these issues in the study of gender gaps in achievement in any setting.

16 For instance, Jacob and Rothstein (2016) document that the issues posed by ordinality may be significantly more severe when examining changes instead of levels. Bond and Lang (2013) provide a quote from Thorndike (1966) which suggests lesser fragility of results to ordinality when considered in levels than in changes: '... it is assumed that the numerals in which the variables are expressed represent equal increments in some attribute. It is also recognised that this assumption is not usually well supported. But for 'rough and ready' studies of relationship, the violation of the assumption does not hurt much. However, when starting to deal with something as fragile as a change score, the violation of this basic assumption becomes a good deal more critical.'

3. The evolution of gender gaps in learning

3.1 Gender gaps in enrolment and grade progression

We first look at gaps in enrolment and grade progression, which are interesting in themselves and important as contextual information to later understand gaps in test scores (Table 1). They also serve to indicate potential magnitudes of selective exclusion that may plague school-based assessments of gender gaps in test scores at different ages.

Table 1. *Enrolment and grade progression by gender and age*

Panel A: Proportion enrolled													
Age	Year	Ethiopia			India			Peru			Vietnam		
		Female	Male	Diff	Female	Male	Diff	Female	Male	Diff	Female	Male	Diff
5	2006	0.04	0.03	0.01	0.45	0.44	0.01	0.01	0.01	0.00	0.01	0.00	0.01*
8	2009	0.78	0.75	0.03	0.99	0.99	0.01	0.99	0.99	0.00	0.99	0.98	0.01
12	2014	0.96	0.93	0.03***	0.97	0.97	0.00	1.00	1.00	0.00	0.98	0.97	0.01
12	2006	0.96	0.94	0.02*	0.87	0.90	0.03	0.99	0.99	0.00	0.96	0.97	0.00
15	2009	0.91	0.88	0.04*	0.74	0.81	0.07***	0.95	0.91	0.03*	0.81	0.73	0.07***
19	2014	0.63	0.56	0.07**	0.42	0.56	0.14***	0.48	0.53	0.05	0.50	0.41	0.09**

Panel B: Highest grade completed													
Age	Year	Ethiopia			India			Peru			Vietnam		
		Female	Male	Diff	Female	Male	Diff	Female	Male	Diff	Female	Male	Diff
8	2009	0.67	0.62	0.05	1.83	1.56	0.27***	1.31	1.32	0.01	1.72	1.70	0.03
12	2014	3.54	3.47	0.07	5.63	5.26	0.36***	6.02	6.02	0.00	5.66	5.61	0.06
12	2006	3.24	3.15	0.10	5.60	5.60	0.00	4.97	4.85	0.12	5.56	5.58	0.02
15	2009	5.69	5.31	0.38***	8.13	8.17	0.04	7.88	7.79	0.09	8.31	8.22	0.09

Notes: This table presents the proportion of students enrolled in formal schooling or higher education at different ages, alongside the average grade level of students (highest completed), by sex. ***p<0.01, **p<0.05, *p<0.1.

At the age of 5, in nearly all countries, children have not yet transitioned to formal schooling; the only exception is India, where about 45 per cent of the sample have joined school. At 8 years old, there are few signs of gender bias in enrolment, with enrolment near universal except in Ethiopia, where a significant proportion of children start schooling later. Rates of grade progression are similar for boys and girls in all countries except India, where girls have on average completed 0.3 grades more.[17]

This absence of gender gaps in enrolment is broadly also true at 12 years old in all countries. Importantly, at 12 years old, enrolment is near universal in all countries. The only exception to this is the Older Cohort in India (where about 10 per cent of children had dropped out by age 12) but even there, the Younger Cohort, seven years later in 2013, has near universal enrolment at the same age. By the age of 15, however, there is a notable gender gap in

17 Perversely, in the Indian context, this is a sign of greater gender bias favouring boys. Specifically, as Singh (2014) shows, children who will eventually enrol in private schooling spend longer in kindergarten classes and start school later. Given that boys are more likely to enrol in private schools than girls, this leads to a higher grade progression for girls in India at younger ages.

enrolment in India (favouring boys) and Vietnam (favouring girls), in both cases about 78 percentage points. There is a modest pro-girl gap in enrolment in Ethiopia at these ages. Finally, by the age of 19, when children are typically in higher secondary grades or college, gender-based differences in enrolment are prominent in all countries except Peru, but differing in direction: enrolment is sharply biased favouring boys in India, but in both Vietnam and Ethiopia favours girls.[18]

3.2 Cross-sectional gaps in test scores

Muted gender gaps in enrolment and grade progression, at least until primary school ages, may still mask significant variation in actual student achievement by gender. Figure 2 presents cross-sectional gaps in tests scores at different ages in the different learning domains, displaying the coefficient on a male dummy from a regression of test scores in each country with 95 per cent confidence intervals; standard errors are clustered at site level in each country. Table 2 presents similar information in tabular form, showing mean differences in the score alongside p-values from t-tests and Kolmogorov-Smirnov tests for equality of distributions.

Table 2. *Test scores of boys and girls from 5-19 years old*

Domain	Age	Cohort	Ethiopia				India				Peru				Vietnam			
			M	F	t-test (p-value)	K-S test (p-value)	M	F	t-test (p-value)	K-S test (p-value)	M	F	t-test (p-value)	K-S test (p-value)	M	F	t-test	K-S test (p-value)
Quantitative skills	5	YC	-0.51	-0.54	0.55	0.60	-0.03	-0.04	0.97	0.81	0.22	0.20	0.63	0.92	0.33	0.32	0.93	0.80
Maths	8	YC	-1.40	-1.42	0.42	0.39	-0.82	-0.84	0.64	0.71	-0.57	-0.64	0.00	0.01	-0.17	-0.16	0.78	0.85
	12	YC	-0.51	-0.55	0.25	0.09	0.06	0.07	0.91	0.36	0.44	0.41	0.29	0.65	0.86	0.92	0.08	0.03
	12	OC	-0.14	-0.16	0.74	0.47	0.33	0.26	0.15	0.04	0.43	0.32	0.03	0.02	0.70	0.74	0.30	0.55
	15	OC	-0.20	-0.41	0.00	0.00	0.15	-0.10	0.00	0.00	0.49	0.51	0.64	0.56	0.80	0.96	0.00	0.00
	19	OC	0.31	0.11	0.00	0.00	0.39	0.11	0.00	0.00	0.77	0.72	0.89	0.05	1.20	1.19	0.84	0.69
Receptive vocabulary	5	YC	0.06	-0.07	0.01	0.00	0.01	-0.02	0.50	0.34	0.03	-0.03	0.23	0.04	0.05	-0.05	0.04	0.11
	8	YC	1.25	1.20	0.44	0.55	0.85	0.68	0.00	0.00	1.71	1.62	0.04	0.38	1.76	1.67	0.03	0.00
	12	YC	2.37	2.29	0.22	0.06	1.96	1.92	0.32	0.49	3.26	3.06	0.00	0.00	2.97	2.99	0.85	0.50
	12	OC	2.45	2.33	0.09	0.08	2.40	2.34	0.39	0.03	2.49	2.37	0.09	0.46	3.46	3.36	0.22	0.39
	15	OC	2.95	2.74	0.01	0.03	2.80	2.38	0.00	0.00	3.86	3.70	0.03	0.01	3.53	3.58	0.59	0.33
Reading	12	YC	-0.04	0.05	0.16	0.26	-0.05	0.05	0.03	0.02	-0.01	0.02	0.63	0.41	-0.14	0.14	0.00	0.00
	19	OC	0.73	0.66	0.41	0.41	0.80	0.51	0.00	0.00	0.87	0.81	0.48	0.88	0.47	0.77	0.00	0.00
Cloze	15	OC	0.03	-0.03	0.41	0.29	0.12	-0.12	0.00	0.00	-0.01	0.01	0.77	0.50	-0.12	0.11	0.00	0.00

Notes: This table presents mean test scores in each country and age group, separately for boys and girls, for each individual test used in this paper. In each subsample, we report p values from t-tests for equality of means and Kolmogorov-Smirnov tests for the equality of distributions. ***p<0.01, **p<0.05, *p<0.1.

3.2.1 *Five and 8 years old*

We find no evidence of a gender gap in mean quantitative skills in any country at 5 years old, that is, at around school entry age; absolute differences between the mean scores of boys and girls are invariably small and we cannot reject equality of distributions in any country.[19] At 8 years old, similar patterns in mathematics hold up; scores remain equal for boys and girls

18 The median enrolled child is enrolled in higher education (university or post-secondary technical) in India, Peru and Vietnam. In Ethiopia, due to a much-delayed age of starting school, the median enrolled child is in late secondary grades.

19 A rejection of the null hypothesis for a two-sided KS test is consistent with first-order stochastic dominance or with the CDFs crossing each other. To distinguish between the two, see the Appendix, where we present CDFs of the test scores for boys and girls for each of the groups in Tables A1 to A6.

across countries, except a small gap in Peru of about 0.07 standard deviations (SD) favouring boys. In receptive vocabulary, these differences are a little larger, albeit still modest in absolute size: at 5 years old, we find small but significant pro-boy differences in Ethiopia and Vietnam and, at 8 years old, continue to find significant gaps favouring boys in three countries, relatively modest in size at ~0.1 SD in Peru and Vietnam, but larger in India at ~0.15 SD.[20]

3.2.2 *Twelve years old*

At 12 years old, with students typically transitioning out of lower primary school (except in Ethiopia), gender gaps in mathematics remain modest in size and typically insignificant. In receptive vocabulary, we see gaps in a similar direction as in mathematics in Peru, about 0.150.2 SD in magnitude. Comparing the two cohorts at 12 years old, to see if gender gaps have importantly shifted between 2006 and 2013, we do not detect any significant differences.[21] On a test of reading comprehension, which was only administered in the Round 4 survey in 2013, we see girls perform better in all contexts but usually with only a small difference, except in Vietnam, where they perform significantly better by about 0.3 SD.

3.2.3 *Fifteen years old*

At 15 years old, when students are transitioning to lower secondary grades or out of school, in contrast with previous ages, gender gaps are substantially larger in magnitude and almost invariably statistically significant. In maths, boys do better by about 0.20.25 SD in Ethiopia and India, while girls do better in Vietnam. This pattern is repeated in the Cloze test of language ability and comprehension administered at this age, with large gaps of about 0.25 SD favouring boys in India and girls in Vietnam. Differences exist also in receptive vocabulary, favouring boys in Ethiopia and India, consistent with the gaps in maths, but also significantly favouring boys in Peru. Comparing the mean differences at 12 and 15 years old for the Older Cohort, this period of early adolescence and the transition out of primary schooling seems a crucial window for the development of gender gaps.

3.2.4 *Nineteen years old*

Finally, at the age of 19, we see gender gaps having crystallised in the direction of gaps seen at 15 years old. We see large gaps favouring boys in Ethiopia and India, and a smaller such gap also favouring boys in Peru. In Vietnam, the country where we are most likely to see gaps favouring girls, mean scores are statistically indistinguishable from each other. By this stage of early adulthood, where a substantial portion of the sample has finished schooling or transitioned out of education entirely, it appears that gender gaps in mathematics do favour boys wherever we find statistically significant gaps, a pattern similar to that documented in international assessments (UNESCO 2015).

20 The modest pro-boy gap at 5 and 8 years old in receptive vocabulary in Vietnam, which otherwise seems to be characterised by either no gaps or pro-girl achievement gaps in both reading and maths, is the one case across all four countries where we see any evidence of the reversal of gender gaps across ages, or heterogeneity across domains of achievement across sex.

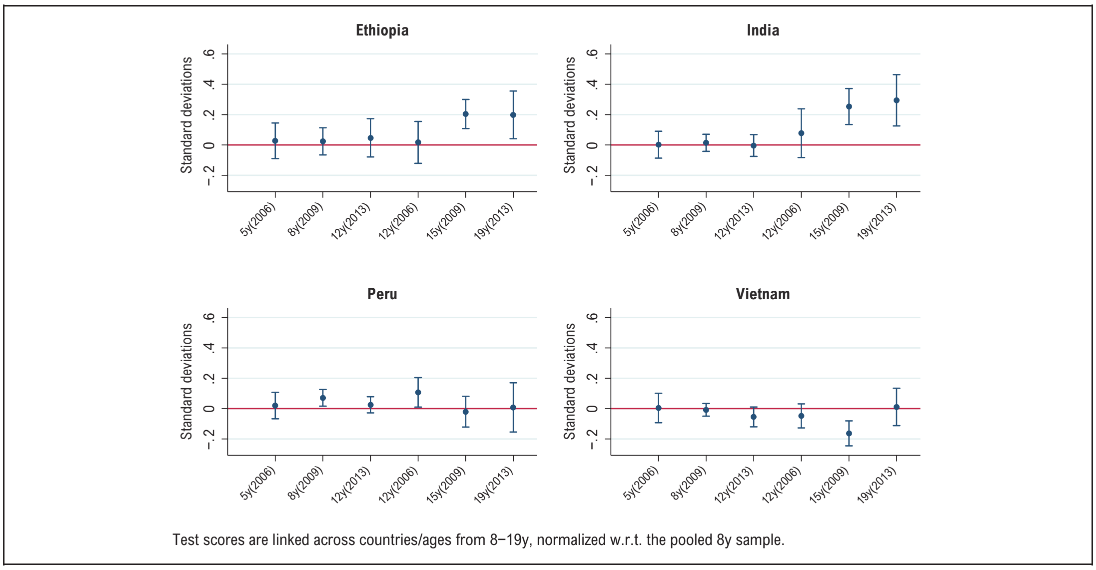
21 Point estimates of the gender gaps at 12 years old are typically similar across the two cohorts for both maths and receptive vocabulary (Figure 2). Note, however, that the confidence intervals are relatively wide.

3.3 Divergence in learning by initial achievement

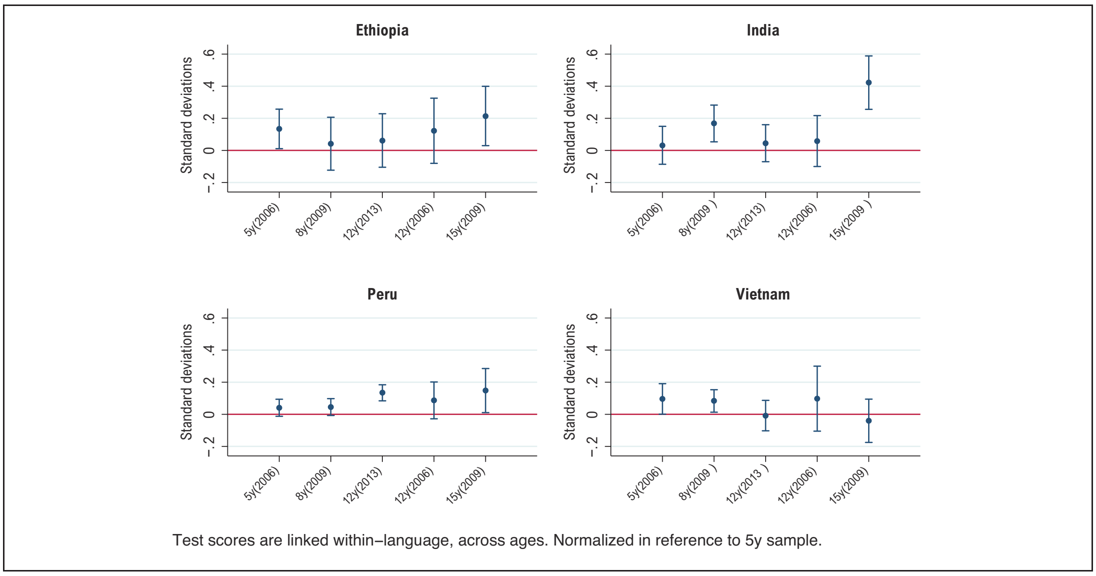
Figure 2 and Table 2, showing the absolute gender gap in test scores, do not answer two important questions. First, to what extent do gaps seen at 15 or 19 years old merely reflect gaps that had arisen earlier but are then perpetuated (and perhaps amplified) by the self-productivity of skills? Second, do these gaps vary across the achievement distribution? For example, do initially well-performing girls continue to progress at par with boys of similar ability but poorly performing girls lag further behind?[22]

Figure 2. *Mean gender differences in achievement from 5-19 years: coefficient plots*

(a) Quantitative skills from 5-19 years old

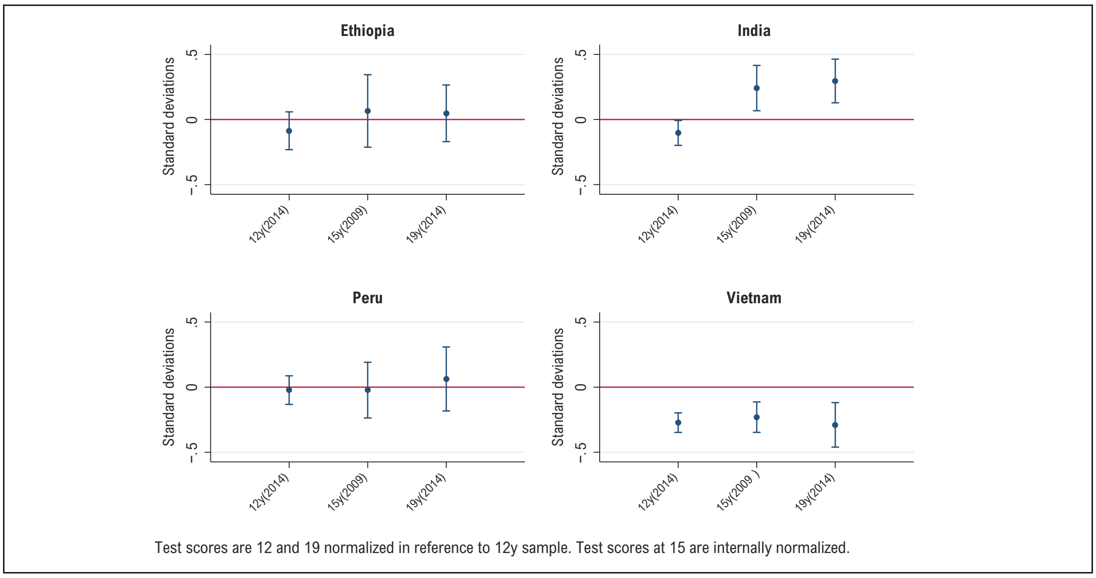


(b) Receptive vocabulary from 5-15 years old



22 Such heterogeneity could result from the process of skill formation directly or reflect different choices made with respect to low performing girls vs. low performing boys, for example in providing remedial investments (such as extra tutoring) to one group and not the other. In either case, understanding the heterogeneity is important for prioritising which groups and age points to target for potential interventions.

(c) Reading scores at 12 and 19 years old



We adopt a straightforward panel-based approach to investigate both these questions. Specifically, we present, for each age group, nonparametric plots which relate current achievement to percentiles of lagged achievement separately for boys and girls.[23] If all of the gap at, say, 15 years old merely reflected the continuation of gaps at 12 years old, then we should see the two nonparametric plots completely overlap each other for a particular country; conditional on having the same baseline achievement, in the absence of fresh divergence between the groups, scores in the next period should also be equal. If there is absolute divergence in test scores for all boys, which is constant across the achievement distribution and does not depend on prior achievement, then we should see the two nonparametric plots be shifted versions of each other with an intercept difference only. Finally, if the divergence depends on initial levels, then we should see a slope difference in the two lines at different levels of prior ability. In case trajectories cross, this would indicate that the direction of gender gaps is reversed depending on whether a child was initially higher or lower performing. Therefore, significant gender gaps could exist for students at a particular point of the distribution, but which cancel out in aggregate and so are not picked up in our previous analysis which only looked at gender gaps in aggregate.[24]

3.3.1 Five to 8 years old

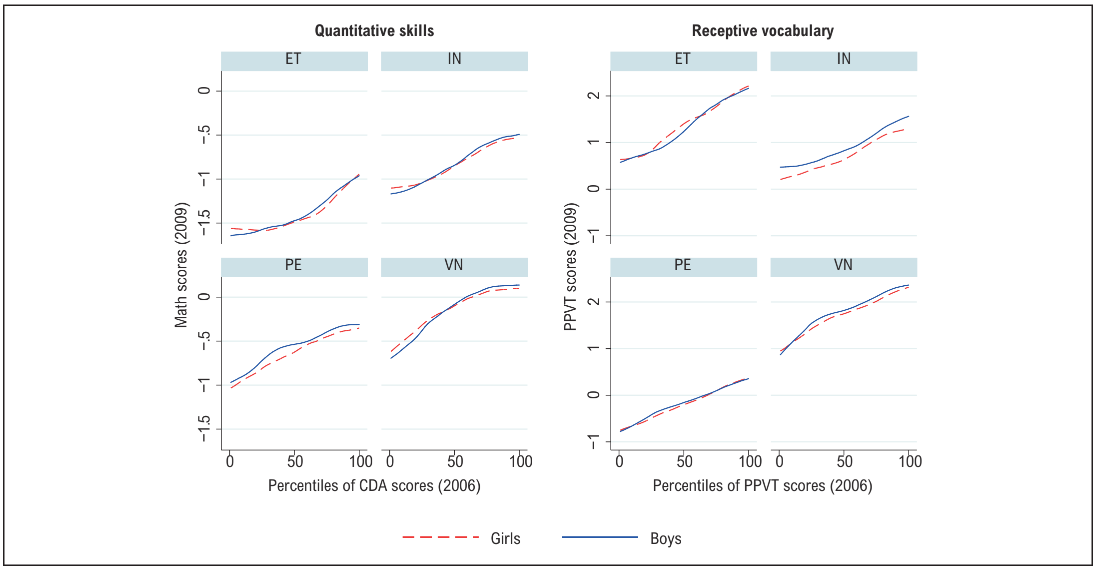
In groups which showed significant differences in test scores at 8 years old, we mostly see the trajectories for boys and girls are different across the full distribution of prior achievement (Figure 3). In particular, note the trajectories for India and Vietnam in receptive vocabulary,

23 These percentiles are defined over the full sample and not separately for boys and girls.

24 As mentioned in Section 2, these graphs also help us consider potential issues arising from ordinality of test scores better. If trajectories do not cross, and the nonparametric curve for one group lie above the other, patterns of divergence in test scores between two ages will be robust to concerns of ordinality in the baseline test scores: at all levels of ability, as one group makes greater progress than the other, any rank-preserving transformation of the prior scores will continue to show this pattern of divergence. If the two curves also have the same slope at different levels of ability, this further implies that the magnitude of the gap will be unchanged regardless of any ordinal transformation.

and for Peru in mathematics.[25] In other groups, where test scores did not differ significantly by gender at 8 years old, we see a near overlap of trajectories of test scores for boys and girls. This indicates that the lack of a gap at 8 years old does represent equivalent progress and not the cancelling out of gender gaps in opposite signs across the achievement distribution.

Figure 3. *Divergence from 5-8 years old*



Notes: Lines are local polynomial smoothed lines (epanechnikov kernel and degree zero) plotted separately for boys and girls.

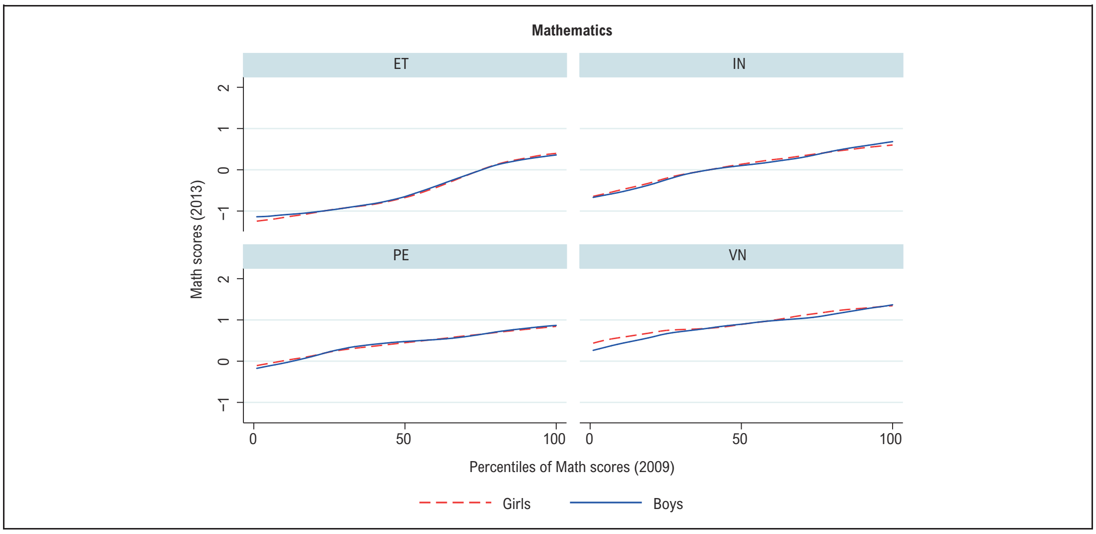
3.3.2 Eight to 12 years old

Trajectories in mathematics seem to overlap near perfectly for all countries (Figure 4), although in Vietnam there are some signs of girls who were low scoring at 8 years old doing better than boys who had scored similarly. The same is largely true of scores on receptive vocabulary. A stark contrast is evident for reading scores where, in all countries, for the range of variation in the lagged vocabulary scores, girls achieve a higher reading score at 12 years old than similarly scoring boys. This difference is very small in Peru but sizable in India and Vietnam, especially for boys and girls in the top half of the achievement distribution at 8 years old.

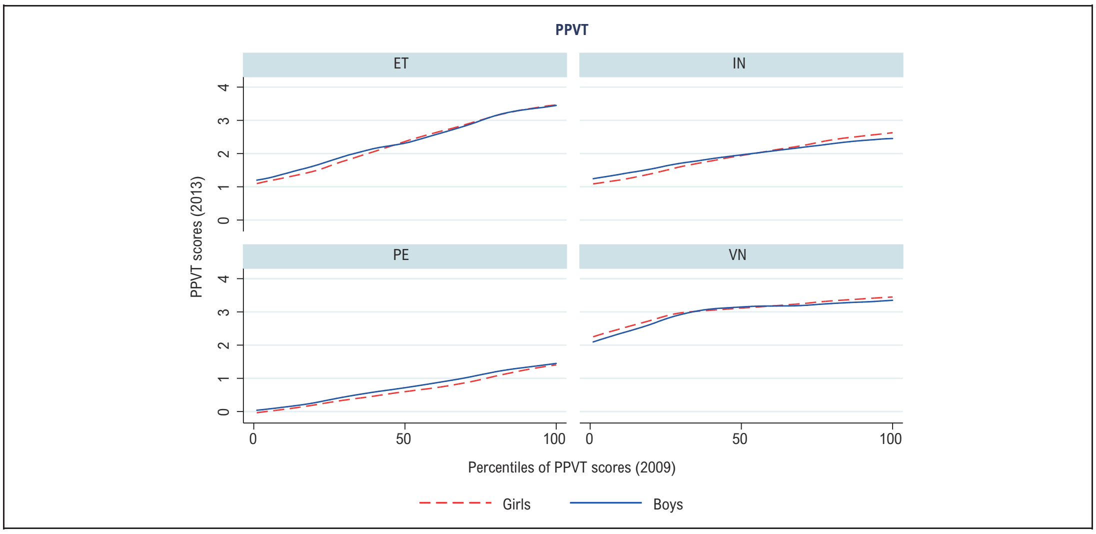
25 The only exception to this pattern seems to be receptive vocabulary in Ethiopia, where we did see significant gender gaps at 8 years old, but where all of this divergence seems to be concentrated in students in the second quartile of baseline achievement.

Figure 4. *Divergence from 8-12 years old*

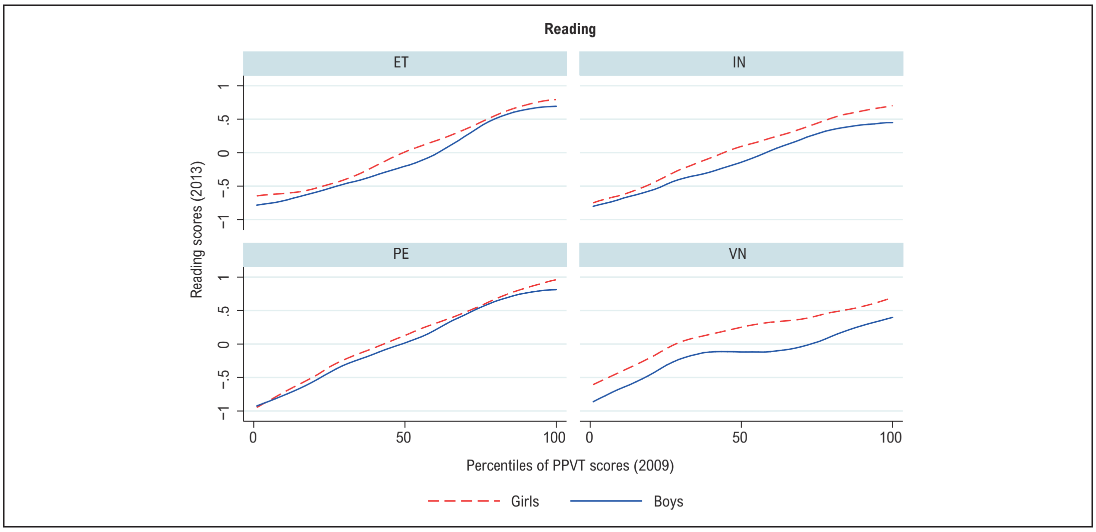
(a) Quantitative skills



(b) Receptive vocabulary



(c) Reading scores

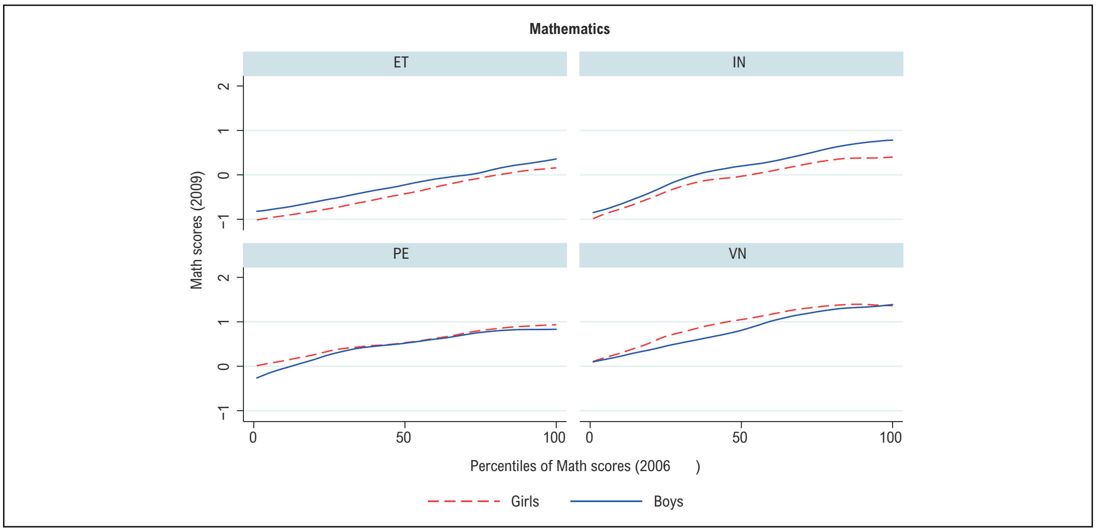


3.3.3 *Twelve to 15 years old*

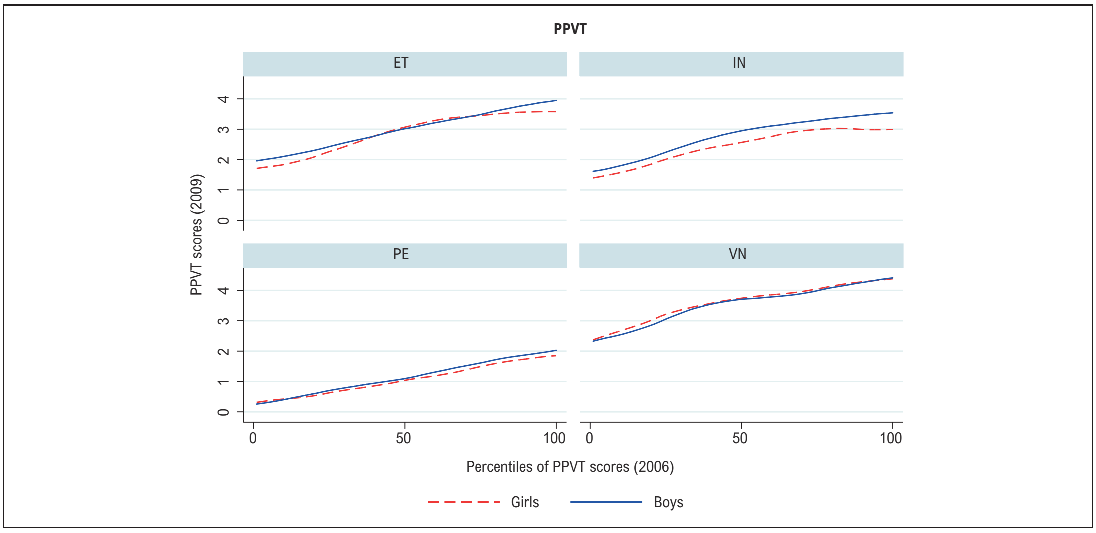
As expected, based on the cross-sectional gaps presented earlier, we see strong evidence of divergence (Figure 5). Ethiopian boys, at all levels of ability at 12 years old, score higher in mathematics than girls at the age of 15. Similar patterns are evident in India, being even clearer, of greater magnitude and evident for all three learning tests administered at 15 years old. In Vietnam, there is clear evidence of girls performing better in reading at 15, regardless of their levels of vocabulary scores at 12; similarly in mathematics, across the bulk of variation in the sample in maths scores at age 12, girls outperform boys.

Figure 5. *Divergence from 12-15 years old*

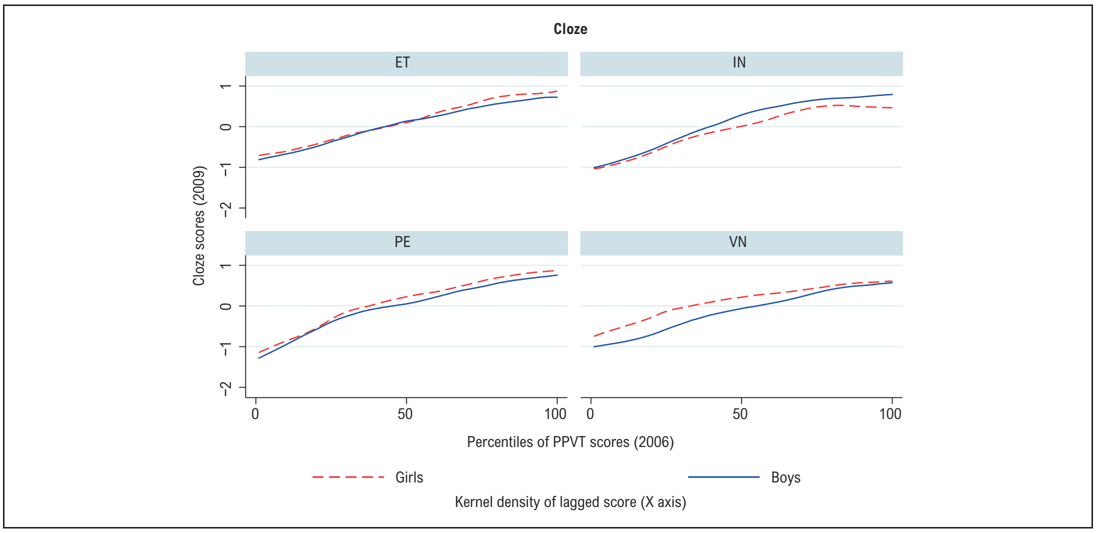
(a) Quantitative skills



(b) Receptive vocabulary



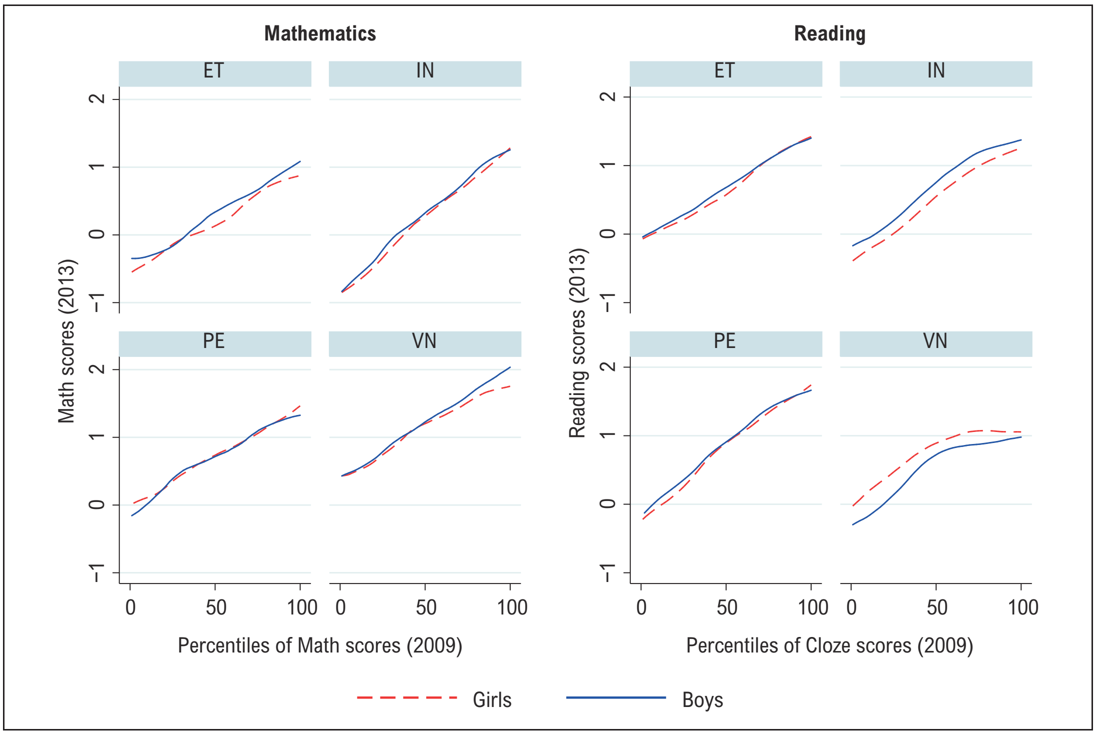
(c) Cloze scores



3.3.4 *Fifteen to 19 years old*

Finally, in the transition to early adulthood, we see a continuation of previous patterns in reading, with modest divergence favouring girls in Vietnam and boys in India at all parts of the initial distribution (Figure 6). In mathematics, the trajectories seem similar for most samples, with at most some modest differences for parts of the achievement distribution.

Figure 6. *Divergence from 15-19 years old*



Notes: Lines are local polynomial smoothed lines (epanechnikov kernel and degree zero) plotted separately for boys and girls.

Summarising over the different age groups considered, we see little evidence of vastly different divergence between boys and girls across the achievement distribution. Where such divergence is evident, it is typically true for all boys or girls and not dependent on initial achievement. Distinct and non-overlapping trajectories of achievement further suggest that such divergence is not an artefact of the ordinal nature of test scores, but rather represents actual difference in the growth of achievement over time.

3.3.5 Imperfect persistence in gender gaps

However, we sometimes see a decline in the absolute magnitude of the gender gap from one age group to the next, even where the panel-based trajectories appear similar for boys and girls. For example, in Vietnam the cross-sectional (pro-girl) gender gap in mathematics scores is clearly evident at 15 years old, but has closed by age 19, even though the rates of learning for boys and girls in Vietnam seem similar.

The explanation for this lies in the empirical regularity of high decay in test scores over time. As a number of studies have documented in different contexts, the persistence coefficient of test scores from one year to the next typically lies between 0.25 to 0.5, even correcting for measurement error leading to attenuation (Jacob et al. 2010; Andrabi et al. 2011).[26] In the case of gender gaps in achievement, this is of particular importance: decay in test scores implies that there may be ‘fresh’ divergence in test scores occurring even if the absolute magnitude of the gender gap measured cross-sectionally remains the same or, in some cases, even declines.[27]

This is of more than academic interest: to the extent that policy priorities are focused towards the reduction of gaps, and in particular to avoiding their exacerbation, panel-based analyses such as ours may indicate age windows for intervention that would not be evident when only looking at the gaps cross-sectionally. Moreover, even analysis of changes in intergroup gaps is not fully informative since effectively it implicitly assumes perfect persistence of learning. Given that this assumption is found routinely violated in panel data, and has important consequences for which ages we think are most necessary to focus on, this is of particular concern in any analyses intending to shed light on the evolution of skill inequalities in childhood.[28]

---

26 We do not correct for measurement error in the analyses above. However, note that under the reasonable assumption that of common measurement error across gender at a given score, any differences in the trajectories of boys and girls are unaltered by potential attenuation bias.

27 Put simply, whether any significant gap declines depends both on decay (which drives gaps towards zero) and the difference in trajectories (which could be in any direction). Where the differences in trajectories are in the same direction as the initial differences in levels, for example, girls scored higher at age 12 and had a higher trajectory between 12 and 15, whether the absolute magnitude of the gender gap increases, stays constant or declines depends on whether the additional gap caused by the higher trajectory exceeds, is the same as, or less than the decline in the gender gap to be expected naturally as a result of decay in test scores.

28 An empirical illustration of this point is the observation in Bond and Lang (2017) that the pattern that a small set of covariates can ‘explain’ the Black-White gap does not mean that later periods of development do not matter in the evolution of inequality in test scores. We are the first, in our knowledge, to stress this in relation to gender gaps in learning.

4. Sources of divergence from 12-15 years old

The rest of this paper focuses on investigating potential sources of the divergence in test scores of boys and girls between 12 and 15 years old. We concentrate on this period since it appears to be a key period for the widening of gender-based gaps in learning as well as enrolment in formal education. We focus on three sets of proximal factors that may individually or jointly explain the gaps: household characteristics and investments into child learning; the time use of children; and differential quality of and experiences in the schools attended.[29]

4.1 Household investments into education

In many domains of human capital in childhood, we know that differences in outcomes result directly from household characteristics or household-level choices regarding investment in children, which makes them an especially pertinent area to investigate potential sources of gender-based divergence.[30]

We look first at a relatively parsimonious set of characteristics and investments. For household characteristics, we use three main variables: caregiver's education, a wealth index, and household size. With regard to investments and child-specific characteristics, we use child-specific information on enrolment, child-specific expenditures on education and children's nutritional status, as summarised by WHO height-for-age z-scores which encapsulate endowments and longer-term investments in health.

Differences by gender with regard to all household characteristics and investments, with the exception of enrolment which was summarised in Table 1, are provided in Table 3. As may be expected if gender is near-randomly distributed, there is not much evidence of differences in the household characteristics of boys and girls in any of the countries. The allocation of investments, however, reveals significant differences. As noted while discussing Table 1, enrolments at 15 and 19 years old displayed significant gender differences in both India and Vietnam. Education expenditures are lower for girls than boys in all countries, except Vietnam at 15 years old, although are only significant in India, where the average annual expenditure on boys' education is double the amount spent on girls. Finally, in nutrition, we do not see many clear patterns of gender gaps in the height-for-age z-scores of children in the four countries at 15 years old, although there is some evidence of a modest gender gap favouring girls in Ethiopia and boys in Peru.

29 We call these proximal factors since they may be caused in turn by more general features such as labour market opportunities in adulthood or social norms not captured here. As recent work has shown (e.g. Jensen (2012) and Munshi and Rosenzweig (2006) in India), changes in these broader economic and social factors may change patterns of differential investments in the education of boys and girls, which may reasonably be expected to affect the inequalities in human capital manifested by differences in test scores.

30 In general, treating the sex of a child as a random event, we would not expect to find systematic differences between the household characteristics of boys and girls. However, if some contexts have selection in sex of the child, for example through sex-selective abortions or selective stopping rules for fertility arising from son preference or gender differences in infant mortality, then systematic differences in household characteristics may still exist. Given that this possibility cannot be ruled out in all our contexts (previous work documents such channels in India), levels differences in household characteristics still merit consideration as a possible channel for gender differences in learning.

The method for the decomposition of gaps is as in the following specifications:

$$Yi,_{15} = \alpha + \beta_1.male_i \quad (1)$$

$$+\beta_2.Yi,_{12} \quad (2)$$

$$+\beta_3.Xi + \beta_4.enroli,_{15} \quad (3)$$

$$+\beta_5.EdExpi,_{15} \quad (4)$$

$$+\beta_6.HAZi,_{15} + \epsilon i,_{15} \quad (5)$$

where Y_{ia} is the current test score for child *i* at age *a*, *male* is a dummy variable (1=male); X_i is a vector of household controls; *enrol* is a dummy variable denoting current enrolment; *EdExp* denotes household spending on the particular child's education; and HAZ_{ia} is the height-for-age z-score; $\epsilon i,_{15}$ is a disturbance term. Regressions are estimated separately for each test at 15 years old for each country, but subscripts for these have been omitted for notational ease.

β_1 is our main coefficient of interest, the interpretation of which changes across specifications. In Equation (1) it is merely the mean difference by gender as in Table 2; in Equation (2), it shows the divergence across gender conditional on past test score, that is, a linear analogue of Figure 3; further specifications investigate whether this divergence can be accounted for by the sequential addition of controls for enrolment and household characteristics (Equation (3)), differences in the educational expenditure for the individual child *i* (Equation (4)), and in nutrition (Equation (5)). Equations (3) and (5), which include lagged achievement, may be considered as a dynamic OLS value-added model (VAM) where the past score provides a summary measure of past investments and individual-specific heterogeneity.

Note that, although parsimonious, the list of investments has good summary measures for household-based investments. Prior achievement should, in a cumulative effects VAM, proxy for past investments. Educational expenditures could summarise a range of different investments into education including, for example, the type of school, extra tutoring, and extra support for buying books and school materials. Similarly, the height-for-age z-scores should be able to proxy for early childhood investments in health and nutrition. Therefore, a priori, we expect that they should be able to account for a substantial share of the variance in learning outcomes.

The goal of our exercise here is not to estimate causal input parameters, but only the extent to which specific channels of factors may explain gender-based differences in test scores.[31] In an ideal setting, we would have experimental variation in each element of the input vector; however, this is not feasible, nor is it possible to find as many valid instruments as the number of inputs. Our decomposition exercise, therefore, relies on a 'selection on observables' assumption.[32] Such an assumption is much more plausible in the dynamic value-added framework than in the cross-sectional decompositions using PISA or TIMSS data in previous work. While it is possible that estimates from VAMs are still biased due to

31 Specifically, we will not engage with whether coefficients from these models should be interpreted as technology parameters or policy effects, an issue that affects many value-added analyses where judgment depends on the full list of variables being controlled for (see Todd and Wolpin (2003); Singh (2015)). We are not interested in the input coefficients per se, but rather in whether these inputs, in unison, can statistically account for the gender gaps that we see in the data.

32 Decomposition exercises, similar to the causal treatment effects literature, need to rely on an assumption of ignorability (conditional exogeneity) to make counterfactual statements. See Fortin et al. (2011) for a detailed discussion.

measurement error and unobserved heterogeneity, in practice the extent of bias seems to be low across a range of applications.[33]

Table 3. *Descriptives of control variables*

	Ethiopia			India			Peru			Vietnam		
	Female	Male	Diff	Female	Male	Diff	Female	Male	Diff	Female	Male	Diff
Household-level variables												
Caregiver's education level												
– None	0.49	0.48	0.01	0.58	0.60	-0.02	0.11	0.12	-0.01	0.08	0.09	-0.01
– Up to Grade 8 (Primary)	0.42	0.45	-0.02	0.27	0.28	-0.00	0.51	0.43	0.08**	0.45	0.46	-0.01
– Grade 9-12 (Secondary)	0.08	0.06	0.02	0.12	0.10	0.03	0.24	0.30	-0.06*	0.40	0.39	0.01
– Above Grade 12 (Post-secondary)	0.01	0.01	0.00	0.03	0.03	0.00	0.15	0.15	-0.01	0.07	0.06	0.01
Household size	6.36	6.34	0.02	5.03	5.08	-0.05	5.34	5.42	-0.08	4.66	4.42	0.24***
Urban	0.43	0.40	0.02	0.24	0.26	-0.02	0.75	0.78	-0.03	0.19	0.21	-0.02
Wealth index	0.35	0.35	-0.01	0.52	0.53	-0.01	0.58	0.59	-0.01	0.63	0.62	0.02
Child-specific investment variables												
Enrolled at 15 years old	0.91	0.88	0.04*	0.74	0.81	-0.07***	0.95	0.91	0.03*	0.81	0.73	0.07***
Annual child specific expenditure on education (nominal, local currency)	130.69	192.81	-62.13	1471.69	3169.24	-1697.55***	295.18	330.17	-34.99	1935.06	1858.73	76.33
Height-for-age z-score	-1.06	-1.78	0.72***	-1.71	-1.62	-0.09	-1.59	-1.37	-0.22***	-1.40	-1.46	0.06

Notes: This table presents mean differences in background characteristics for Older Cohort boys and girls interviewed in 2009. The wealth index is an aggregate of consumer durables, housing quality, and access to services. Height-for-age z-scores are created using WHO reference standards. Annual child-specific expenditures on education include expenditure on fees, extra tuition, uniforms, books and transport. ***p<0.01, **p<0.05, *p<0.1.

Table 4 presents the results from this exercise, with the detailed regression results in Appendix Table B1. The major result, across the different tests and countries, is that these factors jointly explain at most one-third to one-half of the observed cross-sectional gap in test scores at 15 years old. The various controls explain about half of the gender gap in the cloze test in India and about a third in mathematics and receptive vocabulary. In Vietnam, the coefficient on the male dummy declines by about a third in both maths and receptive vocabulary, but in the cloze test is practically unchanged across specifications. In Ethiopia, it appears that the range of controls jointly have near-zero explanatory power regarding the gender gap.

33 In particular, estimates from these models have proven to be unbiased in comparison to experimental estimates (e.g. Deming 2014; Deming et al. 2014; Kane et al. 2013; Angrist et al. 2013; Singh 2015) and to rigorous quasi-experimental estimates (e.g. Chetty et al. 2014; Andrabi et al. 2011; Singh 2014). Note that the ignorability condition does not require all assumptions of the structural cumulative effects model to hold. It requires merely that, conditional on lagged achievement, the inputs are uncorrelated with the error term, a weaker condition. This is akin to the justification behind propensity score matching methods and indeed an essentially similar specification is justified as such, and shown to be unbiased in comparison to lottery-based estimates, by Angrist et al. (2013) in their study of charter schools in the US.

Table 4. *Basic decomposition results: all subjects*

Variables	Country	(1) No controls	(2) + lag	(3) + HH controls + enrolment	(4) + Educ. expenditure	(5) +Height-for-age z-scores
Maths	Ethiopia	0.204***	0.190***	0.206***	0.203***	0.224***
		(0.0458)	(0.0492)	(0.0463)	(0.0459)	(0.0540)
	India	0.253***	0.215***	0.191***	0.169***	0.169***
		(0.0567)	(0.0501)	(0.0455)	(0.0472)	(0.0472)
	Peru	0.0202	0.0797	0.0443	0.0479	0.0471
		(0.0484)	(0.0473)	(0.0397)	(0.0399)	(0.0411)
	Vietnam	0.163***	0.133***	0.112**	0.112**	0.108**
		(0.0393)	(0.0424)	(0.0446)	(0.0448)	(0.0456)
PPVT	Ethiopia	0.212**	0.102	0.147**	0.147**	0.167**
		(0.0896)	(0.0804)	(0.0693)	(0.0694)	(0.0631)
	India	0.422***	0.360***	0.340***	0.324***	0.324***
		(0.0795)	(0.0749)	(0.0745)	(0.0776)	(0.0762)
	Peru	0.148**	0.0659*	0.0690*	0.0668*	0.0584
		(0.0656)	(0.0350)	(0.0376)	(0.0377)	(0.0378)
	Vietnam	0.0404	0.0937	0.0187	0.0189	0.00353
		(0.0644)	(0.0582)	(0.0497)	(0.0491)	(0.0534)
Cloze	Ethiopia	0.0657	0.0805	0.0306	0.0117	0.0117
		(0.131)	(0.119)	(0.112)	(0.135)	(0.135)
	India	0.241***	0.209***	0.147**	0.117*	0.117*
		(0.0831)	(0.0716)	(0.0664)	(0.0642)	(0.0642)
	Peru	0.0228	0.135*	0.120*	0.130*	0.130*
		(0.102)	(0.0734)	(0.0692)	(0.0725)	(0.0725)
	Vietnam	0.230***	0.263***	0.227***	0.221***	0.221***
		(0.0559)	(0.0697)	(0.0664)	(0.0634)	(0.0634)

Notes: Robust standard errors in parentheses, clustered by the initial community the child was surveyed in. See the text for full list of variables included in the regressions. ***p<0.01, **p<0.05, *p<0.1.

Why is this set of relatively rich summary measures affecting achievement still relatively ineffective at explaining gender-based divergence in teenage years? In the individual regressions, covariates typically have expected signs, and are statistically significant for the most important hypothesised factors, indicating that they are relevant for explaining learning even if not the difference in learning for boys and girls. The inability of past achievement and various household characteristics to explain the gender gaps results partly from the fact that gender differences in the factors most predictive of future achievement were relatively small in magnitude in most cases. That we are more successful in explaining gender gaps in India is not because the factors are jointly more predictive – indeed the Rsquare is generally similar across countries – but because it is the only case where lagged achievement, enrolment and child-specific educational expenditures all display statistically significant gender bias in the same direction as the gender gap in learning. Put differently, only in the Indian sample do we see clear differences in those investments underlying achievement production which are directly measured in our data, and therefore it is in India that we can explain a substantial portion of the gender gaps in achievement by 15 years of age.[34]

34 The clearest contrast is with Ethiopia, where the gap in achievement favours boys but the gap in enrolment favours girls. As may be expected, accounting for enrolment actually raises the gap to be explained (columns 2 and 3, Table 4).

Nevertheless, this still leaves open the question about where the gaps in achievement between boys and girls could be arising from, given that they seem not to be arising from the most commonly considered inputs into learning.

4.2 Time use of adolescents

The period from 12 to 15 years of age coincides with adolescence and, potentially, greater involvement in household economic activity and domestic responsibilities. Differences may well open up in how boys and girls allocate their time at these ages and, if not proxied for by previous covariates at the household level, may increase our ability to account for the divergence in test scores.[35] Young Lives surveys collect, for each of the age groups studied in this paper, their time allocation across different uses on a 'typical' day. Such information on time allocation is relatively rare in developing country contexts. This section investigates: (a) whether time allocation patterns differ systematically between boys and girls; (b) whether this is particularly a difference that opens up or widens in the 12-15-years-old period; (c) does the allocation of time across different purposes predict achievement in value-added models?; and (d) does any gender-based difference in time allocation allow us to explain better the widening of gender gaps in this age group?

There is considerable indication of systematic gender differences in time allocation in all countries (Table 5). Girls spend more time on average on domestic tasks and chores, while boys often spend more time working on the family farm or outside the household. These work-related activities become increasingly important after 12 years of age. Looking at direct time inputs into learning, we see that there are some differences in the time spent at school at 12 years old, which line up with the differences in enrolment shown in Table 1. Moreover, there is also a small difference already evident in the time spent studying after school by boys and girls, and that this difference widens by the age of 15; in India, Ethiopia and Vietnam, this is also in the same direction as the gender gaps in learning.[36]

35 Such systematic gender differences may arise for several reasons. For example, gender-specific demands on tasks apart from education may rise more for one group than the other, for example, if girls are expected to contribute more at this age to household chores or if boys are expected to contribute financially to the household requiring dedication of time to paid work. They could arise if social norms are reinforced and internalised which encourage effort by one sex and not the other. Finally, they could arise as adolescents become aware of any gender-based differential in the return to human capital in adulthood and therefore adjust their own effort accordingly (e.g. if better academic results are likely to result in a better job for men but not women). Our focus here is not to separately identify these channels but merely to investigate if accounting for possibly different time use can further explain divergence in learning.

36 The magnitude of the difference is typically about a quarter of an hour a day, going up to about 36 minutes per day extra studying for girls in Vietnam at the age of 15. It is relevant to note that differences in Table 5 do not correct for differential enrolment rates by sex, which is biased in favour of boys in India and girls in Ethiopia, Peru and Vietnam at 15 years of age. In Peru, India and Vietnam, the significantly different enrolment at age 15 likely accounts for at least part of the significant difference by sex in the time spent at school or studying after school: to the extent that this differential allocation is already captured by our controlling for enrolment in the previous section, we do not expect the differential allocation to further explain learning gaps.

Table 5. *Gender differences in time allocation at different ages*

	Ethiopia			India			Peru			Vietnam		
	Female	Male	Diff	Female	Male	Diff	Female	Male	Diff	Female	Male	Diff
8 years old (2009)												
Caring for others	1.07	0.61	0.47***	0.25	0.18	0.07**	0.49	0.47	0.02	0.32	0.17	0.14***
Domestic tasks and chores	2.07	1.29	0.78***	0.44	0.24	0.21***	0.88	0.86	0.02	0.62	0.47	0.15***
Tasks on domestic farm/business	0.79	2.14	1.35***	0.01	0.02	0.01	0.23	0.27	0.04	0.07	0.15	0.08**
Work outside household	0	0.03	0.03*	0.01	0.01	0.01	0	0.01	0	0	0	0
At school	4.91	4.89	0.01	7.61	7.72	0.11*	5.98	5.97	0.02	4.97	4.99	0.02
Studying	0.98	1.01	0.02	1.9	1.77	0.13**	1.9	1.82	0.09*	2.8	2.73	0.07
Play	4.48	4.4	0.08	4.66	4.89	0.23**	4.06	4.3	0.25**	5.48	5.75	0.28***
Sleep	9.7	9.64	0.06	9.11	9.17	0.05	9.67	9.61	0.07	9.72	9.71	0
12 years old (2014)												
Caring for others	0.88	0.45	0.42***	0.19	0.09	0.10***	0.8	0.65	0.15***	0.38	0.3	0.08*
Domestic tasks and chores	2.36	1.27	1.09***	1.01	0.66	0.35***	1.18	1.07	0.10**	1.16	0.88	0.28***
Tasks on domestic farm/business	0.85	2.18	1.33***	0.08	0.16	0.08*	0.43	0.53	0.11*	0.33	0.43	0.10*
Work outside household	0.06	0.1	0.04	0.06	0.04	0.02	0.04	0.04	0.01	0.02	0.04	0.03
At school	5.73	5.49	0.24**	7.96	8.1	0.14	6.14	6.07	0.07	5.47	5.43	0.04
Studying	1.51	1.43	0.09	1.87	1.84	0.03	2.06	1.97	0.09*	2.82	2.61	0.21**
Play	3.39	3.8	0.41***	3.88	4.16	0.28***	3.69	3.82	0.13*	4.89	5.28	0.38***
Sleep	9.22	9.28	0.05	8.95	8.95	0	9.37	9.35	0.02	8.89	9.02	0.14**
12 years old (2006)												
Caring for others	0.69	0.39	0.30***	0.27	0.1	0.17***	0.88	0.6	0.28***	0.33	0.22	0.10*
Domestic tasks and chores	2.83	1.7	1.13***	1.24	0.55	0.69***	1.16	0.98	0.19***	1.29	0.91	0.39***
Tasks on domestic farm/business	0.8	2.04	1.23***	0.2	0.33	0.14	0.32	0.37	0.05	0.56	0.69	0.13
Work outside household	0.12	0.17	0.05	0.4	0.37	0.03	0.03	0.15	0.12**	0.07	0	0.07*
At school	5.74	5.45	0.29**	6.08	6.12	0.04	5.64	5.47	0.18*	4.44	4.41	0.03
Studying	1.74	1.77	0.03	1.83	2.02	0.19*	2.08	1.82	0.26***	2.87	2.59	0.28**
Play	2.75	3.18	0.44***	3.79	4.31	0.53***	2.16	2.32	0.17	5.65	6.11	0.46***
Sleep	9.03	9.02	0.01	9.04	9.04	0	9.29	9.29	0	8.74	9.03	0.29***
15 years old (2009)												
Caring for others	0.91	0.43	0.48***	0.45	0.1	0.35***	0.85	0.66	0.19	0.22	0.11	0.11**
Domestic tasks and chores	3.42	1.69	1.73***	2.05	0.83	1.22***	1.7	1.18	0.52***	1.63	1.25	0.38***
Tasks on domestic farm/business	0.4	2.27	1.86***	0.45	0.54	0.09	0.7	0.65	0.05	0.98	1.13	0.15
Work outside household	0.3	0.48	0.18	1.02	1.05	0.04	0.23	0.58	0.35**	0.4	0.55	0.15
At school	5.75	5.33	0.42**	6.01	6.8	0.79***	6.05	5.77	0.28	4.38	3.99	0.38*
Studying	1.8	1.89	0.08	1.88	2.14	0.26**	2.26	1.94	0.32***	3.27	2.73	0.54***
Play	2.73	3.25	0.52***	3.88	4.25	0.37**	3.09	3.38	0.29**	4.64	5.3	0.66***
Sleep	8.68	8.66	0.01	8.26	8.3	0.03	8.86	8.94	0.08	8.44	8.91	0.47***
19 years old (2014)												
Caring for others	0.97	0.26	0.71***	1.32	0.15	1.17***	2.04	0.37	1.67***	0.82	0.2	0.62***
Domestic tasks and chores	3.18	1.21	1.97***	2.65	1.1	1.55***	2.04	0.99	1.05***	1.8	1.07	0.73***
Tasks on domestic farm/business	0.88	2.46	1.58***	0.96	1.24	0.28	0.78	0.59	0.18	1.07	1.56	0.49**
Work outside household	1.2	2.1	0.91***	1.31	2.9	1.59***	2.21	4.01	1.80***	2.61	3.28	0.67*
At school	3.78	3.43	0.35	3.18	4.25	1.07***	3.32	3.79	0.47	2.92	2.37	0.55**
Studying	1.65	1.58	0.07	1.13	1.24	0.11	1.49	1.47	0.02	1.28	1.04	0.24*
Play	3.74	4.55	0.80***	5.07	5.01	0.07	3.41	3.79	0.38*	5.25	6.14	0.88***
Sleep	8.61	8.42	0.19*	8.37	8.11	0.26***	8.33	8.12	0.21	8.24	8.29	0.05

Notes: Time use was collected based on recall by the respondent of hours spent on the activity on a ‘typical’ day. ***p<0.01, **p<0.05, *p<0.1.

To investigate whether differential time allocation helps account for a larger portion of the divergence in test scores, we follow the specification of Fiorini and Keane (2014) and include a full vector of time use categories.[37] The central message from this exercise (Table 6) is that the information on time allocation of individuals adds little additional to our ability to explain the emergence of gender differences in learning at this age. In no country do we find significant evidence of a decline in the absolute size of the coefficient on the male dummy variable. This is not to say that the information in this vector of time use is as irrelevant as may be seen, both time spent in school and studying have positive coefficients and are frequently statistically significant, but rather that the information was previously already proxied by the controls we had included. In particular, whereas enrolment was consistently positive and (with the exception of receptive vocabulary in Ethiopia) always statistically significant, the inclusion of the time use categories reduces this variable to statistical insignificance in all regressions, indicating that the relevant variation from time use was already proxied by the enrolment variable. Whereas time use differences may have been promising as potential sources of divergence, their additional explanatory power in this case appears minimal.

Table 6. *Do time allocation patterns explain learning divergence across gender?*

Variables	Ethiopia (1) Maths	(2) Vocabulary	(3) Cloze	India (4) Maths	(5) Vocabulary	(6) Cloze	Peru (7) Maths	(8) Vocabulary	(9) Cloze	Vietnam (10) Maths	(11) Vocabulary	(12) Cloze
Male	0.225***	0.302***	-0.129	0.188***	0.337***	0.196***	-0.0229	0.0837**	-0.112*	-0.0906**	0.0521	-0.171***
	(0.0472)	(0.0786)	(0.0855)	(0.0413)	(0.0562)	(0.0604)	(0.0331)	(0.0381)	(0.0605)	(0.0378)	(0.0594)	(0.0571)
Hours per day spent:												
– in caring for household members	0.0360	-0.126**	-0.0464	0.0134	0.0104	0.0225	-0.00580	-0.0273*	-0.00611	-0.0309	-0.0230	-0.0753
	(0.0294)	(0.0555)	(0.0524)	(0.0398)	(0.0604)	(0.0662)	(0.0144)	(0.0158)	(0.0255)	(0.0236)	(0.0515)	(0.0684)
– in household chores	0.0328	0.0596	-0.00552	0.0621**	0.0973**	0.0840**	0.0225	0.0419*	0.0253	0.0300	0.107***	0.0857*
	(0.0240)	(0.0449)	(0.0437)	(0.0264)	(0.0385)	(0.0393)	(0.0212)	(0.0214)	(0.0318)	(0.0254)	(0.0384)	(0.0444)
– in domestic tasks: farming, business	0.0287	-0.0598	0.0500	0.0328	0.0275	-0.0339	0.0182	-0.0121	-0.0113	-0.00224	-0.0194	-0.0163
	(0.0234)	(0.0417)	(0.0412)	(0.0245)	(0.0348)	(0.0365)	(0.0154)	(0.0181)	(0.0264)	(0.0191)	(0.0317)	(0.0327)
– in paid activity	0.0253	-0.0543	0.0268	0.0242	0.0585*	-0.000917	0.0322**	0.00661	0.0133	0.00612	0.00864	-0.0264
	(0.0241)	(0.0499)	(0.0491)	(0.0227)	(0.0322)	(0.0363)	(0.0137)	(0.0151)	(0.0261)	(0.0173)	(0.0289)	(0.0310)
– at school	0.0405*	0.00132	0.137***	0.0752***	0.115***	0.0790**	0.0316**	-0.0211	0.000580	0.0631**	0.0702	0.0632
	(0.0241)	(0.0445)	(0.0426)	(0.0233)	(0.0347)	(0.0362)	(0.0143)	(0.0273)	(0.0368)	(0.0260)	(0.0436)	(0.0416)
– studying outside school	0.121***	0.150***	0.125***	0.0868***	0.125***	0.0865**	0.0540***	0.0548**	0.0643**	0.0187	0.0277	0.0102
	(0.0226)	(0.0431)	(0.0458)	(0.0200)	(0.0319)	(0.0344)	(0.0176)	(0.0217)	(0.0311)	(0.0182)	(0.0293)	(0.0287)
– leisure activities	0.0223	-0.0421	0.0569	0.0388*	0.0724**	0.00426	0.0168	0.00490	0.0268	-0.00661	0.0235	-0.0281
	(0.0213)	(0.0408)	(0.0384)	(0.0216)	(0.0303)	(0.0330)	(0.0130)	(0.0135)	(0.0226)	(0.0190)	(0.0304)	(0.0317)
Lagged maths score	0.392*** (0.0284)			0.440*** (0.0238)			0.424*** (0.0326)			0.429*** (0.0382)		
Lagged vocabulary		0.467***	0.355***		0.406***	0.382***		0.666***	0.760***		0.295***	0.257***
		(0.0371)	(0.0410)		(0.0310)	(0.0323)		(0.0430)	(0.0603)		(0.0339)	(0.0339)
Constant	-1.019***	1.233**	-1.807***	-1.399***	-0.428	-2.055***	-0.312*	-0.0143	-1.188***	0.120	0.949**	-1.223**
	(0.316)	(0.626)	(0.563)	(0.345)	(0.482)	(0.521)	(0.175)	(0.207)	(0.339)	(0.301)	(0.467)	(0.489)
Observations	853	712	534	880	807	774	629	601	604	905	897	888
R-squared	0.41	0.439	0.391	0.575	0.539	0.470	0.504	0.659	0.518	0.5	0.450	0.342

Notes: Robust standard errors in parentheses, clustered by the initial community the child was surveyed in. See the text for full list of variables included in the regressions; coefficients on only key variables are presented in the table. ***p<0.01, **p<0.05, *p<0.1.

37 Since the number of hours in a day total 24, this requires the omission of one category of time use. Here, we choose to omit the number of hours that were spent sleeping. The coefficient on each category of time use therefore should be interpreted as the increment in the productivity of an hour spent in any particular category over an hour spent sleeping.

4.3 Sorting across schools

Thus far, we have focused entirely on household-based measures which might have contributed to gender-based divergence in learning. Schooling, while clearly important to understanding any divergence in learning skills between boys and girls, has been accounted for only indirectly through enrolment, child-specific enrolment expenditures, time spent in school and, possibly, time studying after school (which may be considered to be jointly determined by individuals and schools).

The data, collected through home visits of sample individuals, have limited information on schooling. However, they include unique identifiers for the school attended by the student. We use this information to supplement Equation (5) with a vector of dummy variables for each school attended in the sample.[38] The coefficient on the male dummy variable can thus be interpreted as the remaining gender-based divergence in test scores, conditional on both enrolment and the quality of schools.

Table 7 presents the results. While in no country does school-based sorting succeed in explaining all of the gender-based divergence, the extent to which it can narrow the unexplained portion of the gender-based divergence differs importantly across countries. In Ethiopia, the coefficient on the male dummy looks identical to those in previous specifications, which is consistent with little sorting into particular schools. In India, where gender-based sorting in schools is more of a concern, the coefficient on the male dummy variable declines for all of the tests; however, 40 per cent of the cross-sectional gap in maths scores, half the gap in receptive vocabulary, and a third of the gap in the Cloze test remains unaccounted. Finally, in Vietnam, the coefficient on the male dummy variable for maths declines by half and is now statistically insignificant, and in the cloze test declines by about a fifth, compared to the most extensive specifications shown in Table 4.[39]

Table 7. *Does sorting across schools account for divergence in learning?*

	Ethiopia			India			Vietnam		
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
Variables	Maths	Vocabulary	Cloze	Maths	Vocabulary	Cloze	Maths	Vocabulary	Cloze
Male	0.233***	0.240**	0.0542	0.0705	0.242**	0.0799	0.0400	0.0420	0.177**
	(0.0655)	(0.0934)	(0.132)	(0.0941)	(0.0951)	(0.111)	(0.0475)	(0.0656)	(0.0625)
Lagged maths score	0.336***			0.453***			0.404***		
	(0.0477)			(0.0439)			(0.0510)		
Lagged vocabulary score		0.326***	0.388***		0.568***	0.507***		0.273***	0.274***
		(0.0591)	(0.0408)		(0.0947)	(0.0887)		(0.0712)	(0.0452)
Constant	0.428***	1.685***	0.737***	0.238	1.028***	1.139***	0.362	1.562***	1.132***
	(0.107)	(0.272)	(0.189)	(0.145)	(0.271)	(0.269)	(0.210)	(0.270)	(0.384)
Observations	773	653	486	878	806	772	824	817	808
Rsquared	0.522	0.614	0.585	0.756	0.773	0.691	0.605	0.598	0.466

Notes: Robust standard errors in parentheses, clustered by the initial community the child was surveyed in. See the text for full list of variables included in the regressions; coefficients on only key variables are presented in the table. ***p<0.01, **p<0.05, *p<0.1.

38 We omit the variable for being enrolled; thus non-enrolled students become the reference category and we do not need to omit a school in the regression.

39 We should, however, be cautious in reading too much into the fact that the coefficient is statistically insignificant in some of the tests: the inclusion of a large number of school dummies comes at a cost of statistical power and all estimates are more imprecise than previously.

5. Conclusion

This working paper has focused on the emergence and evolution of gender gaps in learning over an extensive period of childhood, from preschool to early adulthood, in developing countries. The principal contribution is a detailed description of the domains and age periods in which gender-based gaps are observed in four very different contexts. We document that such gaps appear small at primary school ages and grow, particularly in adolescence. Between half and two-thirds of the cross-sectional gaps at 15 years old can be explained with recourse to differences in investments, time use and schooling. However, a substantial unexplained portion remains.

We have focused our analysis on describing gender gaps in achievement and their association with various proximal factors, rather than on estimating causal treatment effects. The results highlight three important areas for policy and further study. The first relates to the timing of divergence: across countries, we find that the period from 12-15 years old is particularly important for the widening of gender gaps in achievement. This implies that policies intended to reduce the eventual gender gap in achievement at the end of schooling should particularly focus on this stage of adolescence/post-primary education. Muralidharan and Prakash (2017) provide an example of such a policy, showing that providing cycles to girls entering secondary schools in Bihar substantially reduced the gender gap in secondary school enrolment in India, which we show here relates to gender gaps in achievement.

The second area which, to us, seems important to note relates to heterogeneity in gender gaps. In one dimension, the domains of achievement, we show that heterogeneity is perhaps less of a concern than in developed countries – where significant, gender gaps in maths and language skills are typically in the same direction. Across contexts, however, heterogeneity in the direction of gender gaps, and their magnitude, seems to be of first order concern. This is important for formulating appropriate policies. While we present evidence for four contexts, it is clear that such analyses would be informative of whether or not gender gaps are a pressing policy concern in any country and how, indeed, they compare with other social and economic inequalities in achievement as claimants for scarce policy attention and resources.[40]

The final area for further study relates to understanding the mechanisms by which gender disparities in achievement emerge. While we have investigated a larger set of such potential channels than previous work, and in a panel setting, a considerable unexplained portion of the gender gap remains, the extent of which differs across countries. This suggests that there is still much room for understanding the mechanisms (and hence potential domains for intervention) in this area. One likely possibility is that some inputs which may contribute to the gender gap are not measured in these data. A further possibility is that the same factors may have heterogeneous effects across sexes or indeed across various background factors (such as sibling sex composition or birth order). We have not investigated such heterogeneity in this paper since, given the limited sample sizes at our disposal, we are not sufficiently powered to investigate this. Studying these potential explanations in appropriate datasets is likely to be a fruitful area for further research.

40 For instance, although we do find occasional gender gaps favouring boys in Peru, the magnitude of these gaps is usually modest and, often, the gaps are transitory at particular ages. In contrast, as we show in a companion paper, gaps relating to socio-economic backgrounds of families are present at all ages, display fresh divergence between all ages and are of significantly larger magnitudes than gender gaps (Krutikova and Singh 2017). In such contexts, prioritising socio-economic gaps for policy and further research may be reasonable.

References

Andrabi, T., J. Das, A.I. Khwaja, and T. Zajonc (2011) 'Do Value Added Estimates Add Value? Accounting for Learning Dynamics', *American Economic Journal: Applied Economics* 3.3: 29–54.

Angrist, J.D., P.A. Pathak, and C.R. Walters (2013) 'Explaining Charter School Effectiveness', *American Economic Journal: Applied Economics* 5.4: 1–27.

Bharadwaj, P., G. De Giorgi, D. Hansen, and C. Neilson (2012) *The Gender Gap in Mathematics: Evidence from Low and Middle Income Countries*, NBER Working Paper 18464, Cambridge, MA: National Bureau of Economic Research.

Bond, T.N., and K. Lang (2013) 'The Evolution of the Black-White Test Score Gap in Grades K–3: The Fragility of Results', *Review of Economics and Statistics* 95.5: 1468–1479.

Bond, T.N., and K. Lang (2017) 'The Black-White Education Scaled Test Score Gap in Grades K-7', *Journal of Human Resources* 0916–8242R.

Chetty, R., J.N. Friedman, and J.E. Rockoff (2014) 'Measuring the Impacts of Teachers I: Evaluating Bias in Teacher Value Added Estimates', *American Economic Review* 104.9: 2593–2632.

Das, J., and T. Zajonc (2010) 'India Shining and Bharat Drowning: Comparing Two Indian States to the Worldwide Distribution in Mathematics Achievement', *Journal of Development Economics* 92.2: 175–187.

Deming, D.J. (2014) 'Using School Choice Lotteries to Test Measures of School Effectiveness', *American Economic Review* 104.5: 406–11.

Deming, D.J., J.S. Hastings, T.J. Kane, and D.O. Staiger (2014) 'School Choice, School Quality, and Postsecondary Attainment', *American Economic Review* 104.3: 991–1013.

Fiorini, M., and M.P. Keane (2014) 'How the Allocation of Children's Time Affects Cognitive and Noncognitive Development', *Journal of Labor Economics* 32.4: 787–836.

Fortin, N., T. Lemieux, and S. Firpo (2011) 'Decomposition Methods in Economics', *Handbook of Labor Economics* 4: 1–102.

Fryer, R.G., and S.D. Levitt (2010) 'An Empirical Analysis of the Gender Gap in Mathematics', *American Economic Journal: Applied Economics* 2.2: 210–40.

Grant, M.J., and J.R. Behrman (2010) 'Gender Gaps in Educational Attainment in Less Developed Countries', *Population and Development Review* 36.1: 71–89.

Guarino, C., M.D. Reckase, and J.M. Wooldridge (2015) 'Can Valueadded Measures of Teacher Performance be Trusted? *Education Finance and Policy* 10.1: 117–156.

Jacob, B., and J. Rothstein (2016) 'The Measurement of Student Ability in Modern Assessment Systems', *The Journal of Economic Perspectives* 30.3: 85–107.

Jacob, B.A., L. Lefgren, and D.P. Sims (2010) 'The Persistence of Teacher-induced Learning', *Journal of Human Resources* 45(4):915–943.

Jensen, R. (2012) 'Do Labor Market Opportunities Affect Young Women's Work and Family Decisions? Experimental Evidence from India', *The Quarterly Journal of Economics* 127.2: 753–792.

Kane, T.J., D.F. McCaffrey, T. Miller, and D.O. Staiger (2013) 'Have We Identified Effective Teachers? Validating Measures of Effective Teaching using Random Assignment', MET Project Research Paper, Seattle: Bill & Melinda Gates Foundation.

Mill, J.S. (1869) *The Subjection of Women, Volume 1*.

Munshi, K., and M.R. Rosenzweig (2006) 'Traditional Institutions Meet the Modern World: Caste, Gender, and Schooling Choice in a Globalizing Economy', *The American Economic Review* 96.4: 1225–1252.

Muralidharan, K., and N. Prakash (2017) 'Cycling to School: Increasing Secondary School Enrollment for Girls in India', *American Economic Journal: Applied Economics* 9.3: 321–50.

Niederle, M. and L. Vesterlund (2010) 'Explaining the Gender Gap in Maths Test Scores: The Role of Competition', *The Journal of Economic Perspectives* 24.2: 129–144.

Nielsen, E.R. (2015) *Achievement Gap Estimates and Deviations From Cardinal Comparability*, Finance and Economics Discussion Series 2015-040. Washington, DC: Board of Governors of the Federal Reserve System

Schoellman, T. (2012) 'Education Quality and Development Accounting', *The Review of Economic Studies* 79.1: 388–417.

Schoellman, T. (2016) 'Early Childhood Human Capital and Development', *American Economic Journal: Macroeconomics* 8.3:145–174.

Singh, A. (2017) *Learning More with Every Year: School Year Productivity and International Learning Divergence*, mimeo. Stockholm School of Economics.

Singh, A. (2015) 'Private School Effects in Urban and Rural India: Panel Estimates at Primary and Secondary School Ages', *Journal of Development Economics* 113: 16–32.

Thorndike, R.L. (1966) 'Intellectual Status and Intellectual Growth', *Journal of Educational Psychology* 57.3: 121.

Todd, P.E., and K.I. Wolpin (2003) 'On the Specification and Estimation of the Production Function for Cognitive Achievement', *The Economic Journal* 113.485: F3–F33.

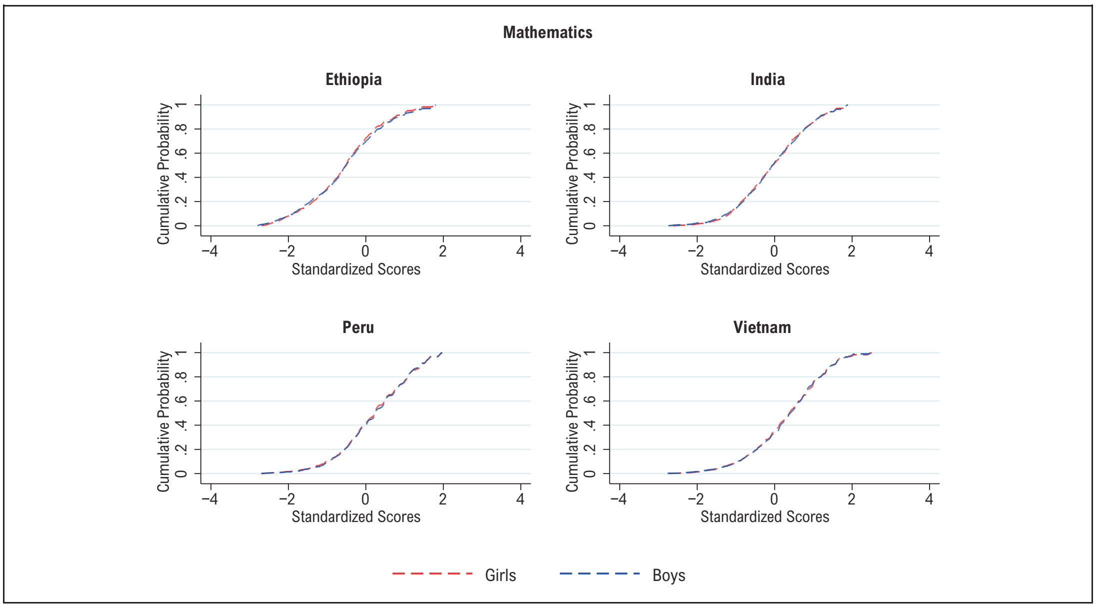
UNESCO (2015) 'Global Monitoring Report 2015: Education for all 2000–2015: Achievements and Challenges', Paris: UNESCO.

Van der Linden, W.J. and Hambleton, R. K. (1997) 'Item response theory: Brief history, common models, and extensions', in W.J. Van der Linden and R.K. Hambleton (eds) *Handbook of Modern Item Response Theory*, 1–28, Amsterdam: Springer Verlag.

Appendix

Figure A1. *CDFs of test scores: 5 years old*

(a) Quantitative skills



(b) Receptive vocabulary

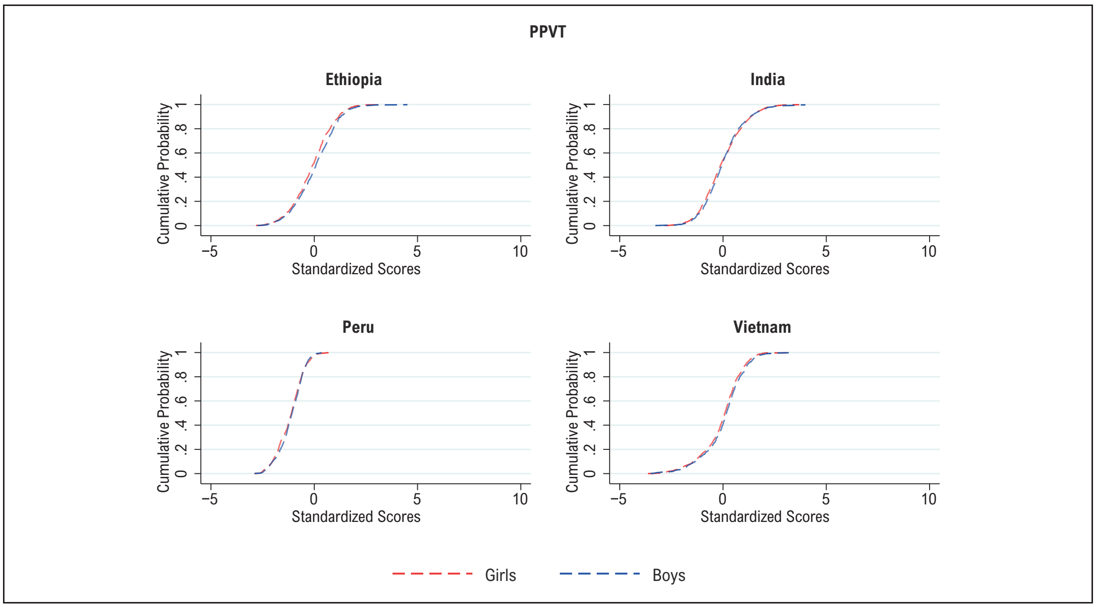
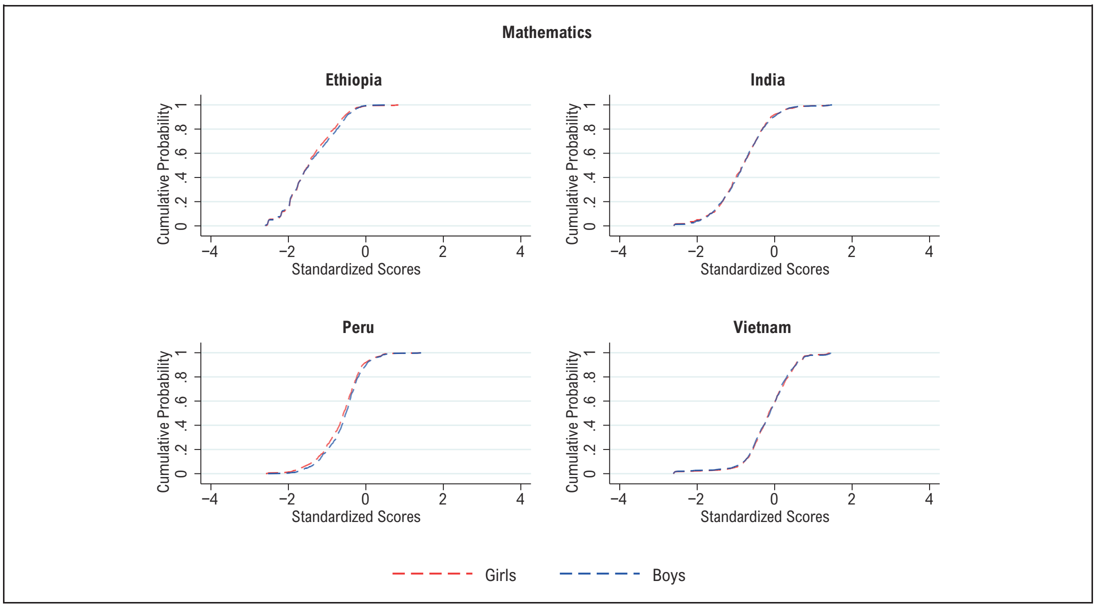


Figure A2. *CDFs of test scores: 8 years old*

(a) Quantitative skills



(b) Receptive vocabulary

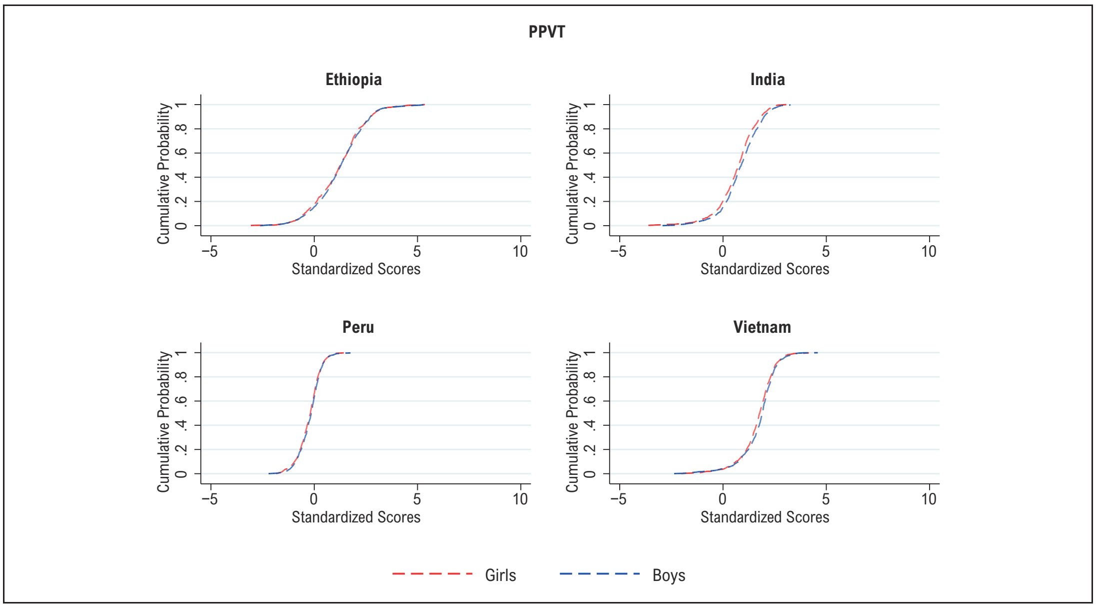
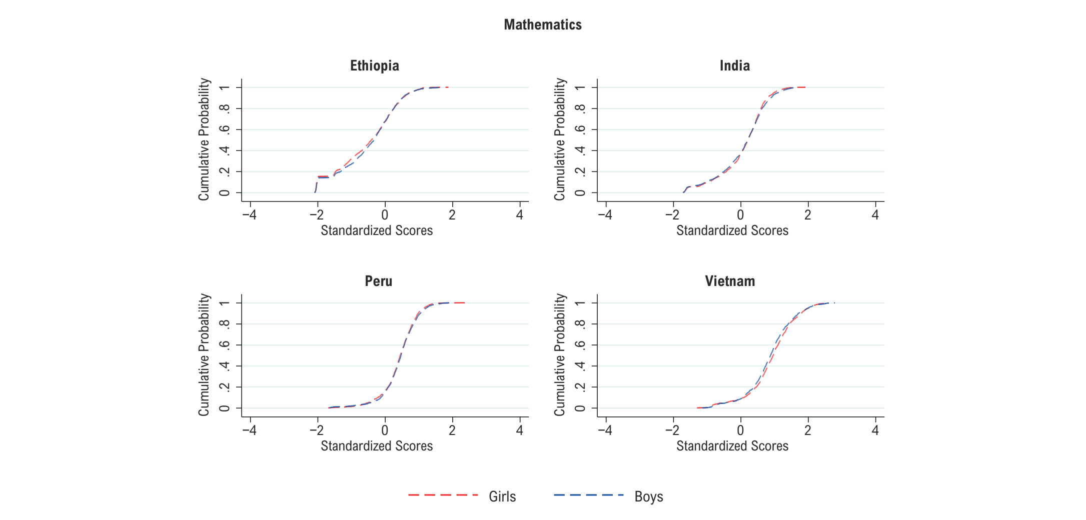
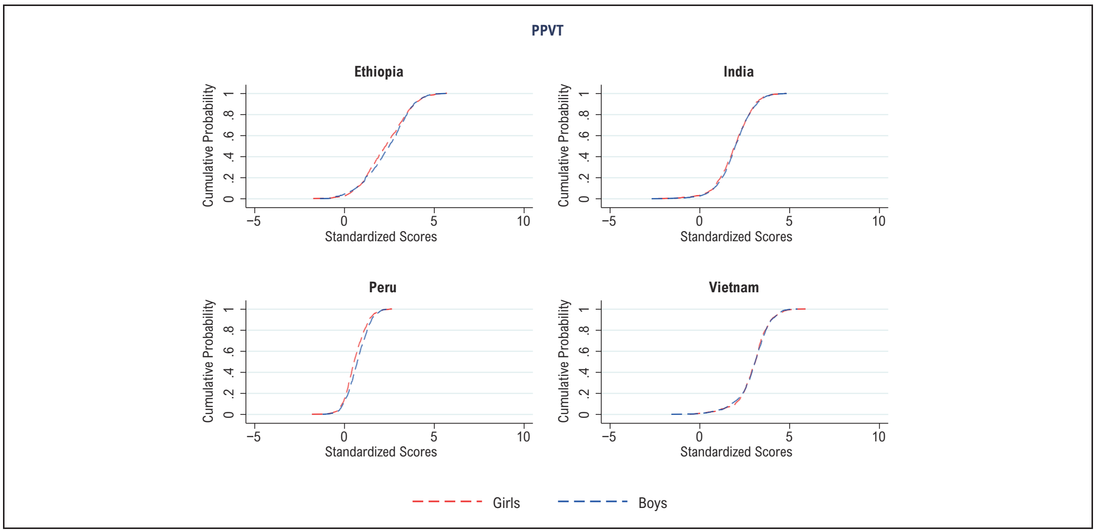


Figure A3. *CDFs of test scores: 12 years old (Younger Cohort, 2013)*

(a) Quantitative skills



(b) Receptive vocabulary



(b) Reading

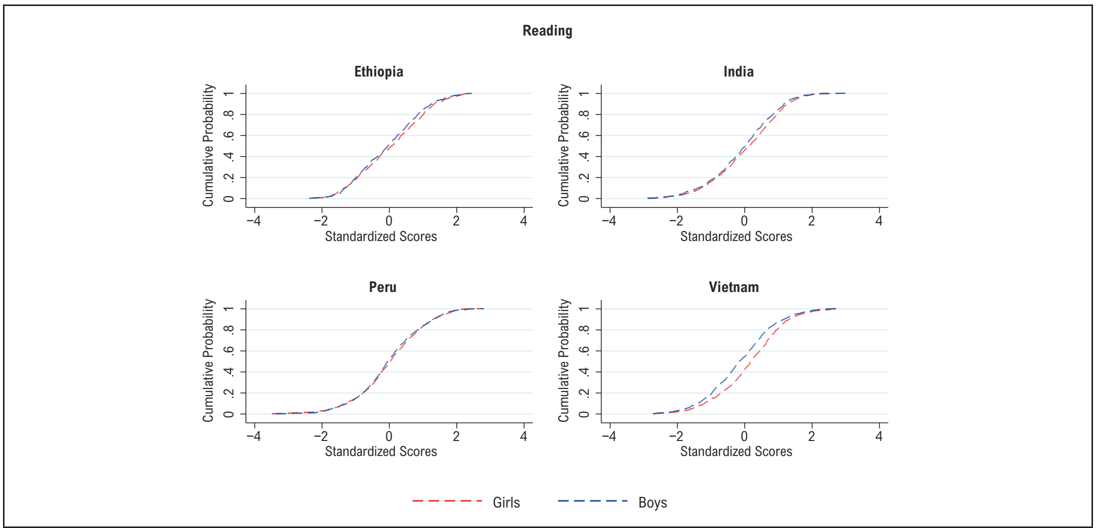
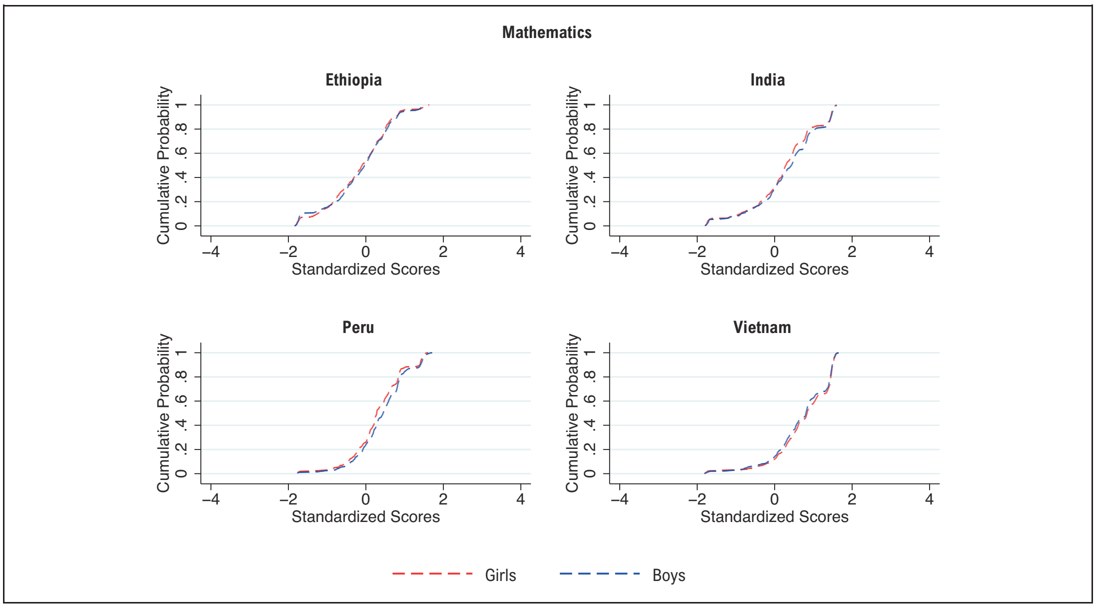


Figure A4. *CDFs of test scores: 12 years old (Older Cohort, 2006)*

(a) Quantitative skills



(b) Receptive vocabulary

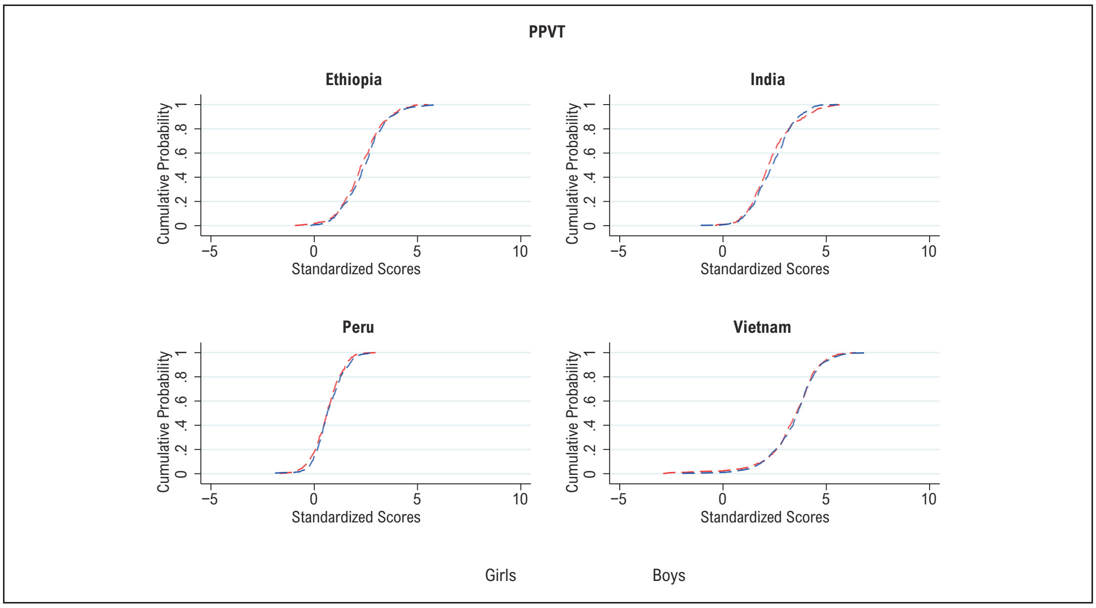
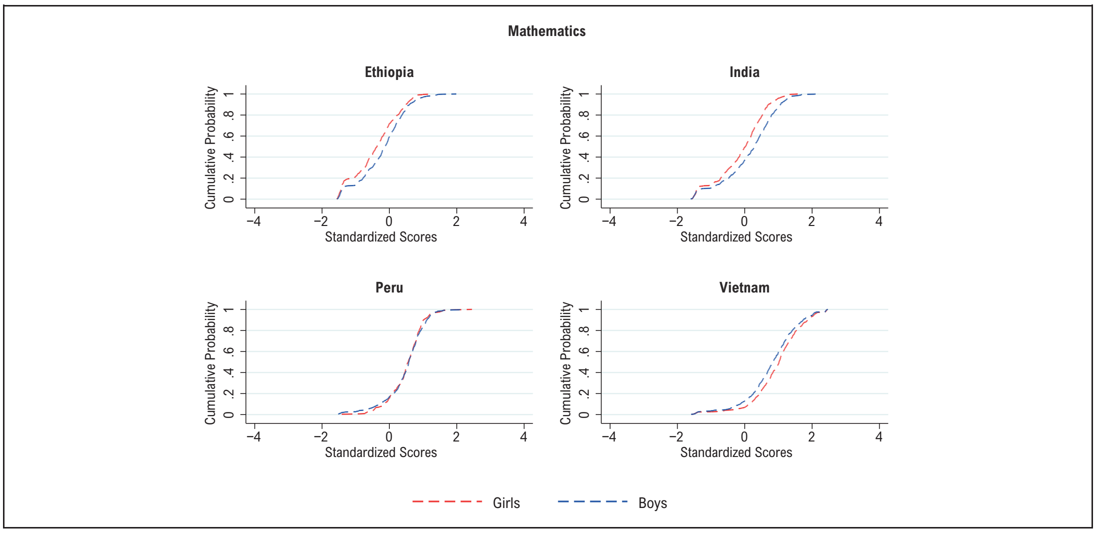
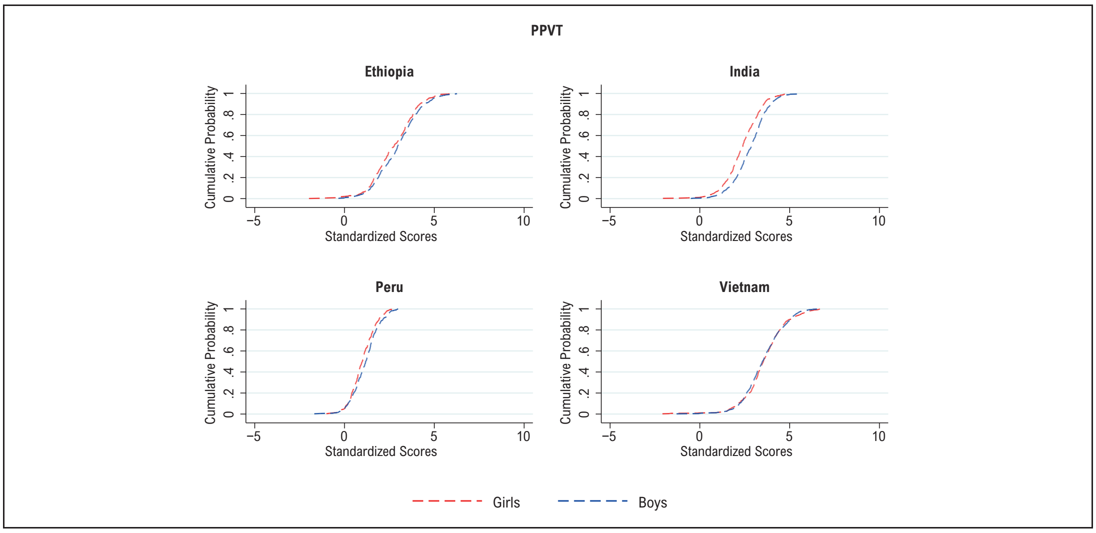


Figure A5. *CDFs of test scores: 15 years old*

(a) Quantitative skills



(b) Receptive vocabulary



(c) Cloze Scores

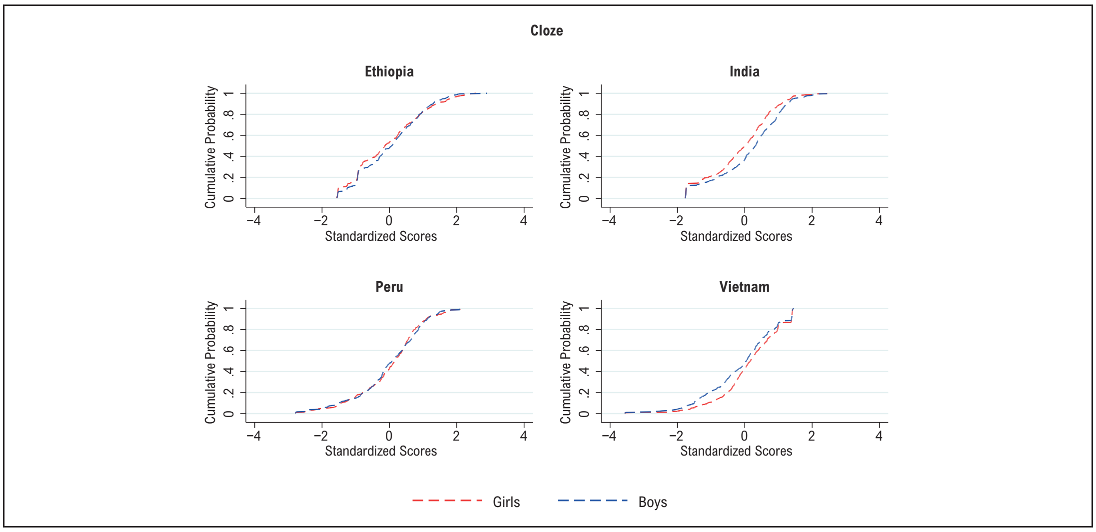
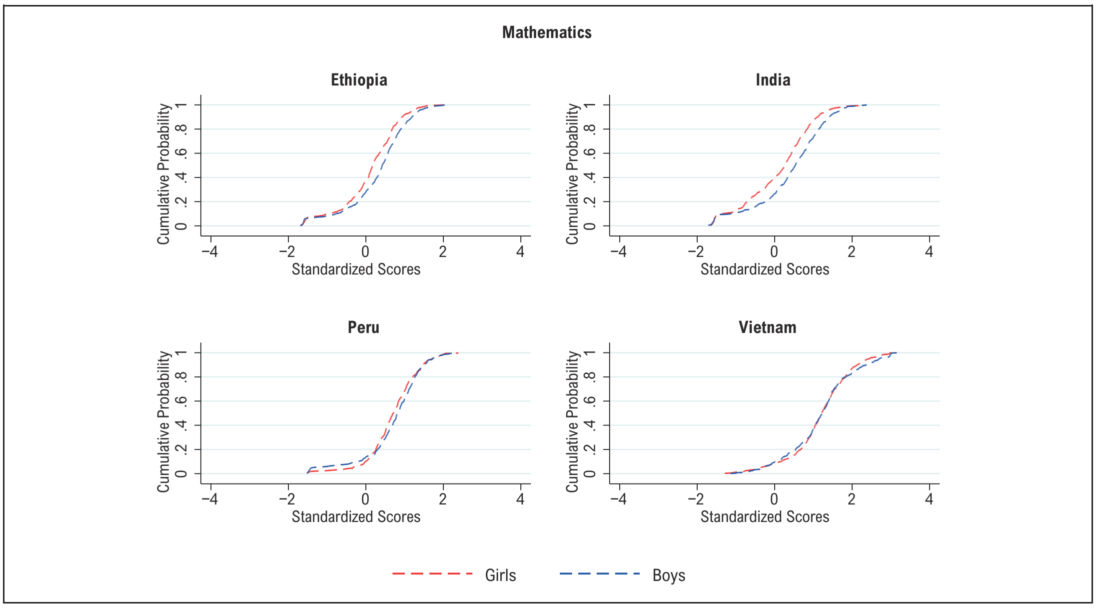


Figure A6. *CDFs of test scores: 19 years old*

(a) Quantitative skills



(b) Reading

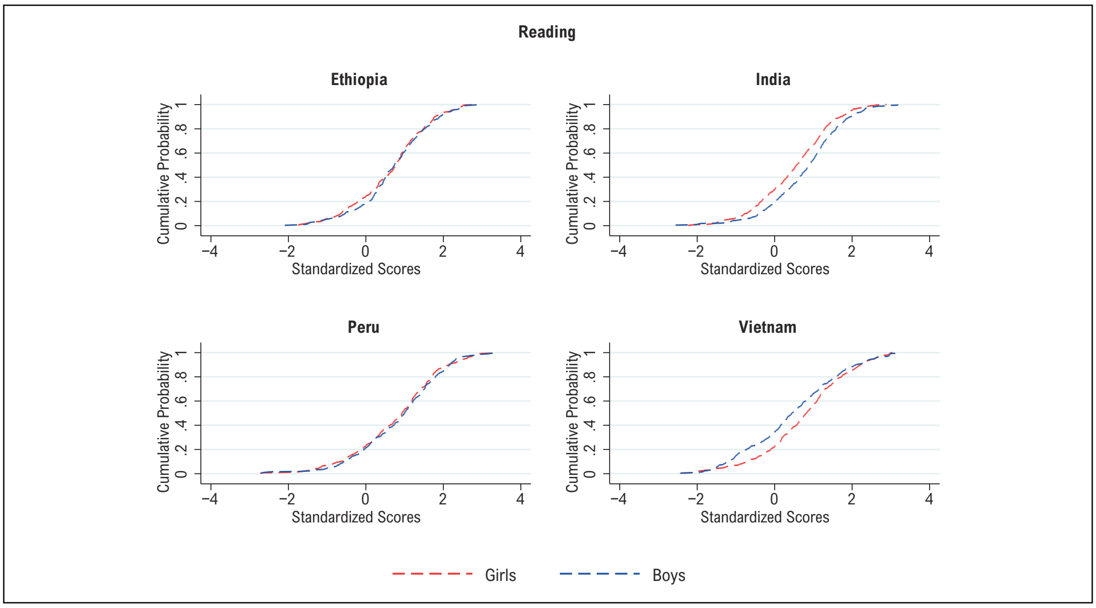


Table B1. *Detailed regression results*

Country	Variables	Maths					PPVT					Cloze				
		(1)	(2)	(3)	(4)	(5)	(1)	(2)	(3)	(4)	(5)	(1)	(2)	(3)	(4)	(5)
Ethiopia	Male	0.204***	0.190***	0.206***	0.203***	0.224***	0.212**	0.102	0.147**	0.147**	0.167**	0.066	-0.081	-0.031	-0.012	-0.012
		(0.0458)	(0.0492)	(0.0463)	(0.0459)	(0.0540)	(0.0896)	(0.0804)	(0.0693)	(0.0694)	(0.0631)	(0.1306)	(0.1187)	(0.1119)	(0.1352)	(0.1352)
	Currently enrolled in school			0.222***	0.220***	0.223***			0.375*	0.375*	0.380**			0.422***	0.429***	0.429***
				(0.0615)	(0.0617)	(0.0611)			(0.1815)	(0.1818)	(0.1795)			(0.0801)	(0.0704)	(0.0704)
	Expenditure on child education				0.048*	0.202***				0.004	0.055				0.025	0.025
					(0.0237)	(0.0520)				(0.0174)	(0.0601)				(0.0651)	(0.0651)
	Height-for-age z-score					0.032					0.020				0.031	0.031
						(0.0237)					(0.0254)				(0.0384)	(0.0384)
	Lagged score		0.503***	0.439***	0.435***	0.417***		0.644***	0.510***	0.509***	0.503***		0.523***	0.396***	0.386***	0.386***
			(0.0317)	(0.0404)	(0.0402)	(0.0407)		(0.0789)	(0.0865)	(0.0875)	(0.0866)		(0.0674)	(0.0439)	(0.0472)	(0.0472)
	Constant	-0.407***	-0.346***	-0.599***	-0.598***	-0.557***	2.743***	1.332***	0.964***	0.965***	1.003***	-0.033	-1.112***	-1.462***	-1.406***	-1.406***
		(0.0723)	(0.0563)	(0.1312)	(0.1309)	(0.1242)	(0.2135)	(0.2308)	(0.3062)	(0.3072)	(0.3018)	(0.1930)	(0.1640)	(0.2331)	(0.2142)	(0.2142)
	Observations	934	933	856	856	854	833	782	715	715	713	633	586	536	534	534
	R-squared	0.022	0.334	0.376	0.378	0.386	0.008	0.322	0.401	0.401	0.398	0.001	0.281	0.357	0.354	0.354
India	Male	0.253***	0.215***	0.191***	0.169***	0.169***	0.422***	0.360***	0.340***	0.324***	0.324***	0.241***	0.209***	0.147**	0.117*	0.117*
		(0.0567)	(0.0501)	(0.0455)	(0.0472)	(0.0472)	(0.0795)	(0.0749)	(0.0745)	(0.0776)	(0.0762)	(0.0831)	(0.0716)	(0.0664)	(0.0642)	(0.0642)
	Currently enrolled in school			0.458***	0.438***	0.437***			0.498***	0.482***	0.483***			0.779***	0.742***	0.742***
				(0.0663)	(0.0646)	(0.0648)			(0.0647)	(0.0672)	(0.0665)			(0.0733)	(0.0688)	(0.0688)
	Expenditure on child education				0.015***	0.015***				0.012	0.011				0.025***	0.025***
					(0.0048)	(0.0048)				(0.0089)	(0.0088)				(0.0077)	(0.0077)
	Height-for-age z-score					0.017					0.054**				0.032	0.032
						(0.0137)					(0.0243)				(0.0260)	(0.0260)
	Lagged score		0.610***	0.467***	0.461***	0.460***		0.586***	0.441***	0.436***	0.434***		0.545***	0.405***	0.395***	0.395***
			(0.0353)	(0.0428)	(0.0430)	(0.0425)		(0.0880)	(0.0845)	(0.0846)	(0.0838)		(0.0787)	(0.0827)	(0.0810)	(0.0810)
	Constant	-0.104	-0.281***	-0.913***	-0.849***	-0.812***	2.379***	1.050***	0.384**	0.447**	0.563***	-0.117	-1.366***	-2.035***	-1.828***	-1.828***
		(0.0639)	(0.0480)	(0.1189)	(0.1159)	(0.1292)	(0.0960)	(0.1636)	(0.1431)	(0.1673)	(0.1670)	(0.0717)	(0.1325)	(0.1167)	(0.1433)	(0.1433)
	Observations	964	964	884	884	884	895	889	812	812	811	859	854	777	777	777
	R-squared	0.027	0.456	0.555	0.561	0.561	0.045	0.417	0.513	0.515	0.517	0.015	0.322	0.436	0.446	0.446
Peru	Male	-0.020	-0.080	-0.044	-0.048	-0.047	0.148**	0.066*	0.069*	0.067*	0.058	-0.023	-0.135*	-0.120*	-0.130*	-0.130*
		(0.0484)	(0.0473)	(0.0397)	(0.0399)	(0.0411)	(0.0656)	(0.0350)	(0.0376)	(0.0377)	(0.0378)	(0.1022)	(0.0734)	(0.0692)	(0.0725)	(0.0725)
	Currently enrolled in school			0.434***	0.426***	0.432***			0.164**	0.158**	0.156**			0.407***	0.407***	0.407***
				(0.0773)	(0.0747)	(0.0768)			(0.0717)	(0.0693)	(0.0708)			(0.1150)	(0.1188)	(0.1188)
	Expenditure on child education				0.109***	0.106**				0.084***	0.082**				-0.090**	-0.090**
					(0.0370)	(0.0382)				(0.0292)	(0.0287)				(0.0370)	(0.0370)
	Height-for-age z-score					0.029					0.029				0.067	0.067
						(0.0193)					(0.0245)				(0.0399)	(0.0399)
	Lagged score		0.517***	0.429***	0.429***	0.433***		0.784***	0.683***	0.676***	0.667***		0.919***	0.775***	0.766***	0.766***
			(0.0338)	(0.0380)	(0.0364)	(0.0317)		(0.0301)	(0.0372)	(0.0357)	(0.0376)		(0.0618)	(0.0780)	(0.0788)	(0.0788)
	Constant	0.510***	0.341***	-0.128	-0.106	-0.044	1.015***	0.544***	-0.047	-0.032	0.043	0.012	-0.517***	-1.217***	-1.079***	-1.079***
		(0.0543)	(0.0287)	(0.1611)	(0.1586)	(0.1565)	(0.1015)	(0.0390)	(0.1195)	(0.1181)	(0.1135)	(0.1380)	(0.0695)	(0.1480)	(0.1890)	(0.1890)
	Observations	673	665	633	633	630	663	636	605	605	602	667	638	608	605	605
	R-squared	0.000	0.364	0.457	0.468	0.491	0.010	0.606	0.644	0.648	0.650	0.000	0.443	0.507	0.512	0.512
Vietnam	Male	-0.163***	-0.133***	-0.112**	-0.112**	-0.108**	-0.040	-0.094	-0.019	-0.019	-0.004	-0.230***	-0.263***	-0.227***	-0.221***	-0.221***
		(0.0393)	(0.0424)	(0.0446)	(0.0448)	(0.0456)	(0.0644)	(0.0582)	(0.0497)	(0.0491)	(0.0534)	(0.0559)	(0.0697)	(0.0664)	(0.0634)	(0.0634)
	Currently enrolled in school			0.406***	0.387***	0.390***			0.358***	0.342***	0.341***			0.472***	0.454***	0.454***
				(0.0557)	(0.0613)	(0.0604)			(0.0961)	(0.1087)	(0.1080)			(0.0699)	(0.0721)	(0.0721)
	Expenditure on child education				0.020	0.019				0.016	0.014				0.021	0.021
					(0.0119)	(0.0120)				(0.0263)	(0.0257)				(0.0164)	(0.0164)
	Height-for-age z-score					0.070**					0.162***				0.068*	0.068*
						(0.0260)					(0.0529)				(0.0338)	(0.0338)
	Lagged score		0.641***	0.459***	0.456***	0.437***		0.544***	0.339***	0.337***	0.310***		0.424***	0.290***	0.274***	0.274***
			(0.0528)	(0.0367)	(0.0354)	(0.0366)		(0.0539)	(0.0533)	(0.0522)	(0.0544)		(0.0326)	(0.0401)	(0.0424)	(0.0424)
	Constant	0.962***	0.483***	-0.007	0.014	0.156	3.575***	1.758***	0.873***	0.893***	1.269***	0.114	-1.332***	-1.569***	-1.389***	-1.389***
		(0.0897)	(0.0635)	(0.2074)	(0.2096)	(0.1959)	(0.1754)	(0.1337)	(0.2658)	(0.2614)	(0.2386)	(0.0973)	(0.1269)	(0.2874)	(0.2680)	(0.2680)
	Observations	968	968	917	917	915	965	960	909	909	907	952	948	900	898	898
	R-squared	0.011	0.379	0.483	0.486	0.492	0.000	0.353	0.421	0.422	0.437	0.013	0.268	0.320	0.326	0.326

Notes: Robust standard errors in parentheses, clustered at the community level. Coefficients on household characteristics are not presented above. See text for the full list of controls.

Starting Together, Growing Apart: Gender Gaps in Learning From Preschool to Adulthood in Four Developing Countries

This working paper studies the evolution of gender gaps in multiple cognitive skills from the ages of 5 to 19 years old, using Young Lives unique panel data from Ethiopia, India, Peru and Vietnam; it is the most extensive panel-based investigation on this question in developing countries.

The findings suggest that, in all four countries, gender gaps in learning are either absent or small in absolute magnitude prior to school entry (at 5 years old) and at primary school age (8 years old). Larger gaps emerge later, widening particularly between the ages of 12 and 15; gaps favour boys in Ethiopia, India and Peru, but girls in Vietnam. This is in contrast to OECD contexts, where significant gender gaps in maths and language skills tend to be in the same direction. Subsequently, these learning gaps appear to mostly persist until early adulthood. In establishing the direction, magnitude, and persistence of gender gaps, we pay careful attention to issues of ordinality and decay in test scores. Panel-based, value-added models with a rich set of covariates including past achievement, child health, time use, parental education and wealth, and school quality, explain at most half to two-thirds of the cross-sectional gender gap in test scores at 15 years old.

Young Lives

An International Study of Childhood Poverty

About Young Lives

Young Lives is an international study of childhood poverty, involving 12,000 children in four countries over 15 years. It is led by a team in the Department of International Development at the University of Oxford in association with research and policy partners in the four study countries: Ethiopia, India, Peru and Vietnam.

Through researching different aspects of children's lives, we seek to improve policies and programmes for children.

Young Lives Partners

Young Lives is coordinated by a small team based at the University of Oxford, led by Professor Jo Boyden.

- *Ethiopian Development Research Institute, Ethiopia*
- *Pankhurst Development Research and Consulting plc, Ethiopia*
- *Centre for Economic and Social Studies, Hyderabad, India*
- *Save the Children India*
- *Sri Padmavathi Mahila Visvavidyalayam (Women's University), Andhra Pradesh, India*
- *Grupo de Análisis para el Desarollo (GRADE), Peru*
- *Instituto de Investigación Nutricional, Peru*
- *Centre for Analysis and Forecasting, Vietnamese Academy of Social Sciences, Vietnam*
- *General Statistics Office, Vietnam*
- *Oxford Department of International Development, University of Oxford, UK*

Contact:
Young Lives
Oxford Department of International Development,
University of Oxford,
3 Mansfield Road,
Oxford OX1 3TB, UK
Tel: +44 (0)1865 281751
Email: younglives@younglives.org.uk
Website: www.younglives.org.uk

Young Lives