Country Report

Young Lives School Survey, 2016–17:
Evidence from Ethiopia

Jack Rossiter, Obiageri Bridget Azubuike and Caine Rolleston

October 2017

Young Lives

Young Lives School Survey, 2016-17: Evidence from Ethiopia

Jack Rossiter, Obiageri Bridget Azubuike and Caine Rolleston

ISBN ISBN 978-1-909403-96-3

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

Jack Rossiter is a Young Lives education research officer working on Ethiopia, and a research associate at the Oxford Department of International Development. His research interests include early learning, the economics of education, cognitive skills development, and educational assessment.

Bridget Azubuike is a quantitative research assistant in the Young Lives education team. She provides support to the research and education teams in the development of the school surveys, data analysis and research dissemination. She has an MSc in Economics from the University of Sussex and a BSc in Economics and Statistics from University of Benin, Nigeria.

Caine Rolleston is a senior lecturer at the Institute of Education at University College, London and senior education associate at Young Lives. His research interests focus on educational access, learning metrics, educational effectiveness and the economic benefits of education.

Acknowledgements

The authors would like to thank: Professor Tassew Woldehanna, Chanie Ejigu, Meseret Gebresilassie, Yeshi Mulatu and Solomon Areaya at the Ethiopia Development Research Institute; Alula Pankhurst, Young Lives Ethiopia Country Director; all the survey supervisors and enumerators during 2016-17; the Ethiopian Ministry of Education for support and leadership; the Ministry of Education's Maths and Science Improvement Centre and Educational Initiatives for assessment design and development; and Rhiannon Moore and Padmini Iyer for their comments and contributions to this report.

Most of all, we thank the schools, teachers and students who participated in the survey.

About Young Lives

Young Lives is an international study of childhood poverty, following the lives of 12,000 children in 4 countries (Ethiopia, India, Peru and Vietnam) over 15 years. **www.younglives.org.uk**

Young Lives is funded by UK aid from the Department for International Development (DFID).

The views expressed are those of the author(s). They are not necessarily those of, or endorsed by, Young Lives, the University of Oxford, DFID or other funders.

Summary

This report gives an overview of the Young Lives school effectiveness survey conducted at the beginning and end of the 2016-17 (2009 E.C.) academic year. It provides a descriptive summary of the data collected from 12,182 students in Grade 7 and Grade 8, learning in 30 sites across seven of Ethiopia's eleven regional states and city administrations.

The school effectiveness survey was designed to allow analysis of what shapes children's learning and progression over a school year. The data will allow researchers to understand, describe and explain school and education system effectiveness. The survey focuses on issues of attainment (e.g. progression and grade completion) and achievement (e.g. on the learning and other outcomes delivered by the school system) in a structure that links students to teachers, classrooms and schools.

1. Introduction and overview

1.1 The Young Lives study

Young Lives is a unique longitudinal study of childhood poverty which has been conducted in Ethiopia, India, (Andhra Pradesh and Telangana), Peru and Vietnam since 2002. The study traces the lives of 3,000 children in two age cohorts (a 'Younger Cohort', born in 2001-02 and an 'Older Cohort', born in 1994-95) in 20 sentinel sites located in each country. In Ethiopia, sites are located in five regions: Addis Ababa, Amhara, Oromia, Southern Nations, Nationalities and Peoples' Region (SNNP), and Tigray.

Data are collected on a variety of human development topics, including health, nutrition, education and livelihoods. The structure of the survey allows the comparison of the same children at different ages, to see how their lives are changing; and of different children at the same age, to see how their communities have changed over time. The Young Lives research programme combines periodic data collection at the household and community levels (carried out in 2002, 2006, 2009, 2013 and 2016) with longitudinal qualitative research and nested school surveys.

In 2010, a school component was introduced to explore Young Lives children's experiences of schooling and education in depth. Primary school surveys were conducted in India (2010), Ethiopia (2010), Peru (2011), Vietnam (2011-12) and again in Ethiopia (2012-13). In 2016-17, a further round of Young Lives school surveys was conducted at the upper primary level (in Ethiopia) and secondary level (in India, Peru and Vietnam).[1] This report covers the 2016-17 Young Lives school survey in Ethiopia.

Young Lives' survey data are all publicly archived with the UK Data Service (http://ukdataservice.ac.uk). Data from the school survey will be added to the archive by June 2018.[2]

1 We targeted the grade with the highest share of 14-15 year olds in each country – equivalent to our Younger Cohort children. Different school starting ages and system structures leads to the variation in grades selected across countries.

2 For further details see www.younglives.org.uk/content/use-our-data.

1.2 The school survey in Ethiopia

The 2016-17 Young Lives school survey in Ethiopia (the 'school survey') allows researchers to understand, describe and explain school and education system effectiveness using observational data. The survey focuses on issues of attainment (e.g. progression and grade completion) and achievement (e.g. on the learning and other outcomes delivered by the school system). Priority areas for upper primary education policy were identified through consultation with the Government of Ethiopia's Ministry of Education, development partners and associated education stakeholders. These priority areas guide the main research questions:

- At what level are students performing in core curricular and transferable domains[3] and are levels indicative of preparedness for further education and training?
- How much progress are students making in one academic year and what are the drivers of learning trajectories over time, including how these relate to equity?
- What is the role of key dimensions of education quality in shaping educational outcomes over time and, in particular, which teacher practices are associated with improved learning outcomes?
- What are the relationships between language of instruction (intended and applied), participation, learning levels and preparedness for further education and training in secondary grades?

1.3 Sample design and site selection

The first Ethiopia school survey (2010) followed Young Lives Younger Cohort children into schools to gather information about their educational experiences. The sample for the second Ethiopia school survey (2012-13) included Young Lives' 20 core sentinel sites (in Addis Ababa, Amhara, Oromia, SNNP and Tigray) and was extended to include 10 sites in the 'emerging' regions of Somali and Afar.[4]

The 20 core sites were selected purposively in 2001 to ensure that the household survey reflected the cultural and geographic diversity of the country, including urban–rural differences, but with a pro-poor bias and a focus on areas with food insecurity (Outes-Leon and Sanchez 2008). Between three and five sites were selected in each region.

The 10 sites in Somali and Afar were added for the school survey because in these emerging regions, historically poor access to and use of services is of interest to government and education stakeholders. Sites were selected in these regions according to the same criteria as the initial 20 sites, but with additional considerations for fieldworker safety and security.

The school survey included the same 30 sites as in 2012-13. In each of the 30 sites, a census was conducted, including all schools within the site's geographic boundary, irrespective of ownership. Included in data collection at these schools were: all Grade 7 and Grade 8 students; all teachers of mathematics and English of Grade 7 and Grade 8 students;

3 'Transferable domains' are widely understood to include a range of cognitive and non-cognitive skills that provide young people with 'critically needed tools to be able to succeed in terms of employment, health and personal well-being' (see Rankin et al. 2015).

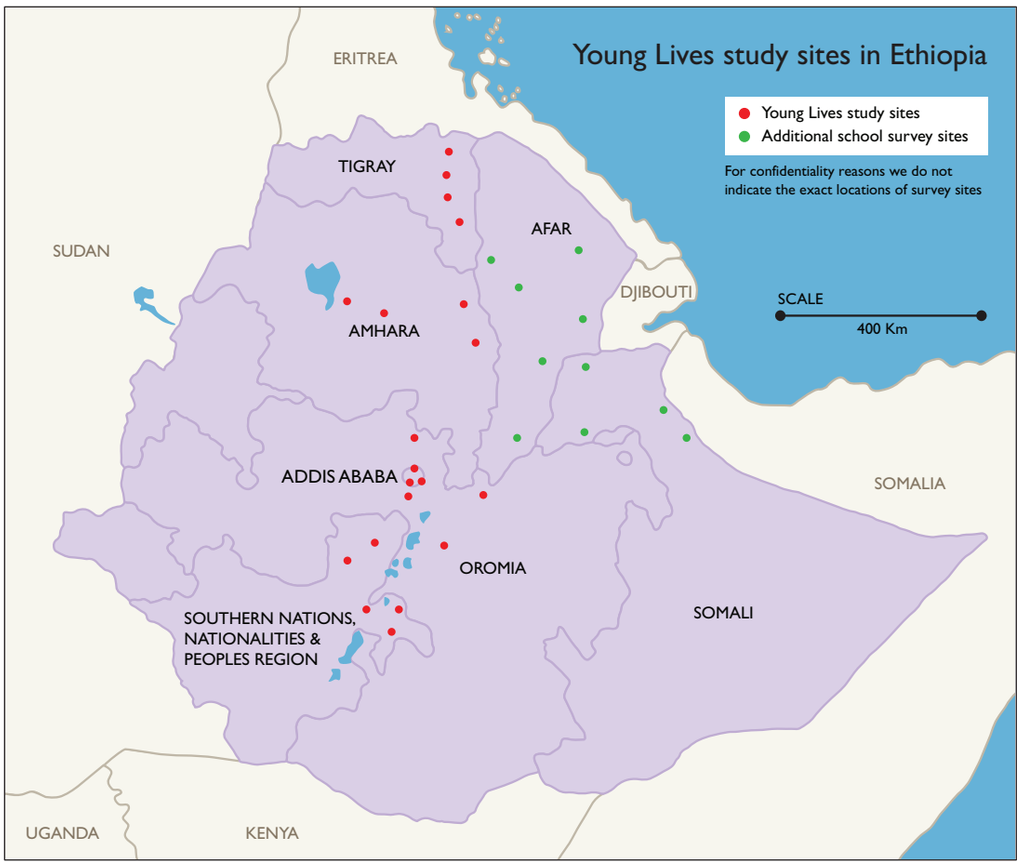
4 Four regional states, Afar, Somali, Gambella and Benishangul-Gumuz, are known as 'Emerging National Regional States' because they have low levels of development relative to the national average. Recognising this, the Government of Ethiopia has established a multi-sectoral special support programme, with a substantial education component.

and all school directors. This census included both Young Lives Younger Cohort children and non-Young Lives children.

The survey was conducted at the beginning (Wave 1) and end (Wave 2) of the school year in collaboration with the Ethiopian Development Research Institute (EDRI). At Wave 1, the student sample included all students present on the first day of student tests in each school. This set of students was then followed up at Wave 2, without replacement of absentees.

This sampling approach permits multiple links to be made between the students and schools in the second and third survey rounds (i.e. between 2012-13 and 2016-17), including comparisons of attainment and achievement at different stages in the education system. The sample is statistically representative at the level of the site and allows in-depth investigation of learning levels, gains and the different factors associated with progress for children from different backgrounds and settings. Appendix 1 provides a description of the 30 school-survey sites, and Figure 1.0 shows the geographic distribution of survey sites.

Figure 1.0. *Map of school survey sites*



1.4 Survey design and content

Survey instruments administered included student, teacher, section- and school-level questionnaires and observation protocols. Figure 1.1 provides an overview of instruments used.

At Wave 1, Grade 7 and Grade 8 students' learning levels were measured using two cognitive tests: mathematics and functional English ('English'). Questionnaires were used to collect data on students' and teachers' backgrounds. Each school director completed a set of questions about himself/herself and his/her school. In addition, fieldworkers completed an observation of the school facilities and collected administrative data on student attendance.

At Wave 2, Grade 7 and Grade 8 students' learning levels were re-measured using two cognitive tests: mathematics and English (see Box 1.0). These students also completed a questionnaire of psychosocial scales (e.g. school engagement and motivation), provided their perceptions of the 'classroom instructional environment', and completed a third cognitive test of functional Amharic ('Amharic'). Teachers completed a questionnaire and selected psychosocial scales relating to teacher motivation. Mathematics teachers also completed a 'teacher professional knowledge' questionnaire, designed to measure their specialised content knowledge for teaching. Finally, fieldworkers collected administrative data on student and teacher absenteeism, retention and change during the school year.

Figure 1.1. *Survey instruments administered*

Student outcome measures	Background instruments
Mathematics test Repeated measures, administered at the beginning and end of the year. Assessing students' curriculum knowledge and ability to apply that knowledge in less familiar contexts.	**Director questionnaire** Collects background data on the director and the school.
Functional English test ('English') Repeated measures, administered at the beginning and end of the year. Assessing students' English reading and comprehension skills relevant to the contexts in which they may use the language.	**Teacher questionnaire** Collects background data on teachers, including teacher motivation and section-level information.
Functional Amharic test ('Amharic') Single measure, administered at the end of the year. Assessing students' Amharic reading and comprehension skills relevant to the contexts in which they may use the language.	**Student questionnaire** Collects background data on students (including academic support within and beyond school, psychosocial measures and perceptions of the classroom instructional environment).
	School facilities observation Collects data on school infrastructure and facilities.
	Teacher professional knowledge questionnaire Collects mathematics teacher performance on an assessment of specialised content knowledge for teaching.

Box 1.0. *The 'value-added' approach*

Often, students, teachers and schools are ranked by achievement *levels* (for example, in the Ethiopian National Learning Assessment, which provides a check of levels against expected learning criteria (Ministry of Education 2013)). The problem with this approach, however, is that achievement is tightly linked to a student's socio-economic status and his or her learning to date. In an observational dataset, with non-random sorting of students into schools and classrooms, there is a substantial risk that measures of learning are systematically correlated with measures of school and classroom inputs and processes. As a result, when attempting to understand the relationships between student performance and these inputs and processes, estimates are likely to be biased.

The goal of the value-added approach is to allow 'apples-to-apples' comparisons of how much teachers and schools contribute to student progress each year. This means focusing not on how students test at a single point in time, but rather on how much improvement they make from one testing period to the next. Because the value-added approach adjusts for students' prior performance and observable background characteristics, one school or teacher value-added estimate can be compared with another without fear that the estimate is heavily biased by selection processes or other variables. The value-added approach therefore returns a less-biased estimate of how much teachers and schools add to the students who learn in their schools – with strong research and policy implications (for an interesting example of the value-added approach in use, see Kane and Staiger 2008).

1.5 Test and instrument development

Cognitive test instruments were developed in 2015-16 through a process of item development, selection, translation and adaptation, and qualitative and quantitative pilot testing. The following sections provide an overview of the tests administered in the school survey.[5]

1.5.1 Mathematics test

In the 2016-17 school survey, we conceptualise learning quality both in terms of progress on curriculum knowledge *and* students' ability to apply their knowledge and skills in less familiar contexts.[6] This is to reflect the priorities of education systems at upper primary and secondary levels. We identified the TIMSS[7] mathematics assessment framework as a useful starting point for assessing students' mathematical ability as it distinguishes between the following three cognitive domains:

- *Knowing*: the facts, concepts and procedures students need to know.
- *Applying*: the ability of students to apply knowledge and the conceptual understanding to solve problems or answer questions.
- *Reasoning*: going beyond the solution of routine problems to encompass unfamiliar situations, complex contexts and multi-step problems (Grønmo et al. 2015).

5 See Azubuike et al. (2017) for more details on the design and development of these tests.

6 See Iyer and Moore (2017) for a more detailed discussion of the way in which quality learning has been conceptualised in the 2016-17 school surveys.

7 The Trends in International Mathematics and Science Study (TIMSS) is an international assessment of mathematics and science at Grades 4 and 8 that has been conducted every four years since 1995. See: http://timss2015.org.

We worked with the Ministry of Education's Mathematics and Science Improvement Centre (MSIC) to select content domains for the mathematics test. Experts from this centre are engaged in a multi-year process to develop an item bank for formative and summative assessments for students in Grade 7 and Grade 8. Eight content domains (Table 1.0) were identified, based on Ethiopia's minimum learning competencies in Grades 5 to 8. In addition, the desire to create a linked test across India, Vietnam and Ethiopia (the three countries in which Young Lives was implementing a school survey with a value-added design in 2016-17) meant that those other two countries' curricula influenced the selection of common domains.

Table 1.0. *Content domains for mathematics*

Basic number competency	Geometry and shapes
Integers, rational numbers, powers and bases	Algebra
Fractions, decimals, ratios and percentages	Measurement, charts and graphs
Area, perimeter, volume and surface area	Reasoning, problem solving and applications in daily life

Based on these content domains, 149 mathematics test items were developed with the MSIC. These, along with items from Educational Initiatives'[8] item bank, served as a pool from which 117 items were selected for piloting across three forms.[9] Items were mapped according to content and cognitive domain, grade level and relevant minimum learning competency. Further details of the piloting procedures for all instruments are in Part 4.

1.5.2 English test

English language tests were included in the school survey as a reflection of the status of English in Ethiopia as a 'transferable skill', with relevance for continuing education, labour market opportunities and social mobility (Graddol 2010). In Ethiopia, English serves as a language of instruction for all secondary-level subjects (excluding Ethiopian languages).

The 'Functional English' construct has been defined as the 'application of [English] skills in purposeful contexts and scenarios that reflect real-life situations' (OFQUAL 2011). In this sense, the test diverges somewhat from the school curriculum in Grade 7 and Grade 8. Instead, it covers content that students at upper primary grades should be familiar with after learning English as a second language from Grade 1 and will need for secondary school. Due to practical considerations of conducting a large-scale survey in schools, the test was paper-based and focuses on the following four skill domains:

- *Word identification*: identifying simple vocabulary which students are likely to have been exposed to, with particular focus on language relating to their everyday environment and to education.
- *Word meaning and contextual vocabulary*: identifying the meaning of less-familiar words through their contextualised use in a sentence, or via a synonym/antonym.

8 Educational Initiatives conduct large-scale assessments and education programmes throughout India and internationally, working with state and national governments and non-governmental organisations. Young Lives has worked in partnership with Educational Initiatives on the development of the mathematics and English tests in the 2016-17 school surveys in Ethiopia, India and Vietnam.

9 Criteria for pilot item selection included: content and cognitive domain (to ensure relevance and coverage), grade level, and minimum learning competency (to indicate difficulty and coverage). In addition, to aid translation and minimise the interference of literacy/language with the assessment of mathematics, we favoured items with a short and concise stem and simple vocabulary.

- *Sentence construction and comprehension:* completing sentences correctly, using appropriate grammatical concepts and/or combining sentences.
- *Reading and comprehension:* reading a range of texts (stories, posters, factual passages) and comprehending direct facts and implicit inferences.

Pilot items were selected from Educational Initiatives' large item-pool, based on contextual suitability. In total, 90 items were piloted across two test forms. Items were mapped according to content domain, skill domain and approximate Common European Framework of References for Languages (CEFR) level.[10] Further details of the piloting procedures for all instruments are in Part 4.

1.5.3 Amharic test

Amharic is Ethiopia's official federal working language and while it is not the Ethiopian language with the largest number of mother tongue speakers (that is Afaan Oromo) it does serve as a lingua franca for business, commerce and government. Among Ethiopia's 91 registered linguistic groups, Amharic is the dominant language of communication and influences social mobility, rural–urban migration and social integration. Amharic proficiency strengthens students' economic participation and labour market prospects because, in particular, a large share of job opportunities in major towns and cities becomes accessible (Alemu et al. 2011; Vujcich 2013). Amharic is taught from Grade 3 to all students for whom it is not their mother tongue and it is expected that variation in proficiency will relate to language of instruction models applied in different regions, and student pathways from mother tongue to second and third languages.

Following consultation with the English Language and Mother Tongue Directorate at the Ministry of Education, the Amharic test was included in the second wave of data collection. To maintain consistency with the English cognitive test, the CEFR served as the basis for the test of Amharic skills. A pilot set of 36 items was prepared, covering *basic*, *independent* and *proficient* users and including the types of Amharic used in the social, academic and professional contexts experienced by students in upper primary grades and expected in the few years after completing primary schooling. It focused on the everyday expressions and phrases used in interaction with other Amharic speakers; it did not seek to understand a student's grasp of Amharic literature or of complex Amharic grammar that might be taught to a mother-tongue speaker. Further details of the piloting procedures for all instruments are in Part 4.

1.5.4 Final cognitive test specification

Pilot data were analysed to generate a range of statistics for item selection. Techniques from Classical Test Theory (CTT) and Item Response Theory (IRT)[11] were used, focusing on face validity, construct validity and internal consistency reliability at the test level and difficulty, discrimination and the quality of distractors at the item level. The test development and item selection process returned five tests (Table 1.1).

10 The CEFR framework details six levels of English language proficiency. Young Lives English test items were graded according to the types of language skills and proficiency they required.

11 See Azubuike et al. (2017) for a more detailed discussion of the process of pilot data analysis and item selection in the mathematics and English tests.

Table 1.1. *Final five cognitive tests*

	1	2	3	4	5
Subject	Mathematics	Mathematics	English	English	Amharic
Wave	Wave 1	Wave 2	Wave 1	Wave 2	Wave 2
Linkage across waves	Linked between Waves 1 and 2		Linked between Waves 1 and 2		n/a
Students assessed	All		All		Two sections per school
Number of items	40	40	40	40	18
Format	Multiple-choice				

1.5.5 *Student background questionnaire*

The student background questionnaire was completed by all students in Grade 7 and Grade 8 at each school included in the survey sample. It was administered across the two waves and contained three parts:

- Part 1: home and family background, educational history and experiences, life outside school.
- Part 2: psychosocial scales covering future orientation, effort, personal development and self-efficacy. Scales were selected following piloting and because of their expected association with students' academic outcomes.[12]
- Part 3: student perceptions of the 'classroom instructional environment', once in relation to their regular mathematics lessons and once for their regular English lessons.[13]

1.5.6 *Director background questionnaire*

The school director background questionnaire was completed at each school included in the survey sample. It was administered at Wave 1 and contained two parts:

- Part 1: director background information, professional training and experience, duties at the school.
- Part 2: school background information, funding, personnel and performance in regional assessments.

1.5.7 *Teacher background questionnaire*

The teacher background questionnaire was completed by all mathematics teachers and English teachers of Grade 7 and Grade 8 at each school included in the survey sample. It was administered across the two waves and contained two parts:

- Part 1: teacher background information, professional training and experience, approaches to educational assessment and duties at the school.
- Part 2: teacher attitudes and beliefs, including psychosocial scales on wellbeing, relationship with others in the school, and efficacy.[14]

12 See Little and Azubuike (2017) for further details.

13 See Moore and Rossiter (forthcoming) for more details on the measure.

14 See Moore and Rossiter (forthcoming) for more details on the measure.

1.5.8 Teacher professional knowledge questionnaire

A measure of mathematics teacher professional knowledge (TPK) was completed by all mathematics teachers of Grade 7 and Grade 8 in schools included in the survey sample. Items were adapted from the Learning Mathematics for Teaching project[15] and covered three content domains: geometry; numbers, concepts and operations; and patterns, functions and algebra.[16]

2. Survey findings

2.1 Description of the final sample

The final sample, presented in Table 2.0, included 63 schools and 271 sections (134 at Grade 7 and 137 at Grade 8). The majority of surveyed schools are government-owned (51 out of 63), but 12 non-government-owned schools were also surveyed in sites in Addis Ababa, and the SNNP and Somali regions.[17]

Table 2.0. *School and section sample, by location*

Location	Number of sentinel sites	Schools			Sections		
		Government-owned	Non-government-owned	Total	Grade 7	Grade 8	Total
Addis Ababa sites	3	7	2	9	23	27	50
Amhara sites	4	7	0	7	15	14	29
Oromia sites	4	6	0	6	11	12	23
SNNP sites	5	9	3	12	27	29	56
Tigray sites	4	7	0	7	20	17	37
Somali sites	4	7	7	14	25	25	50
Afar sites	6	8	0	8	13	13	26
Total	**30**	**51**	**12**	**63**	**134**	**137**	**271**

Table 2.1 details the student-level sample and instrument completion at Wave 1 and Wave 2. A total of 13,943 students were recorded as enrolled (in school registers) at the time of Wave 1, of which 12,182 (87 per cent) were attending during the Wave 1 survey visit. These 12,182 students constitute the survey sample. Of these survey students, 11,735 were recorded as still enrolled in school at the time of Wave 2, but only 10,735 (88 per cent of those attending at Wave 1) were present during the Wave 2 survey visit. A total of 8,855 students completed all survey instruments in both waves.

15 The Learning Mathematics for Teaching Project (LMT) investigates the mathematical knowledge needed for teaching using measures that reflect the real mathematics tasks teachers face in classrooms, for instance, assessing student work, representing numbers and operations and explaining common mathematical rules or procedures. See: www.umich.edu/~lmtweb.

16 For more details of the items and adaptation, see Moore and Rossiter (forthcoming).

17 The inclusion of non-government schools in certain sites must be borne in mind in interpreting results. Non-government schools include 'private' schools, 'faith' schools and 'community' schools, according to the major source of finance.

Table 2.1. *Student sample and instrument completion, by location and survey wave*

Location	Wave 1			Wave 2		Waves 1 and 2
	All students enrolled at time of W1 [A]	Students included in survey [B]	Survey students completed all W1 instruments [C]	Survey students still enrolled at W2	Survey students present at W2 [D]	Survey students completed all W1 and W2 instruments
Addis Ababa sites	2,458	2,302	2,244	2,233	2,057	1,913
Amhara sites	1,288	1,185	1,129	1,128	1,029	745
Oromia sites	1,194	1,041	989	992	954	820
SNNP sites	2,838	2,406	2,204	2,284	2,128	1,669
Tigray sites	1,857	1,707	1,643	1,671	1,575	1,310
Somali sites	3,114	2,571	2,387	2,510	2,181	1,680
Afar sites	1,194	970	947	917	811	718
Total	**13,943**	**12,182**	**11,543**	**11,735**	**10,735**	**8,855**

Notes: [A] All students listed as being enrolled on the school register. [B] Included in survey if attending on first day on which student instruments were administered at Wave 1. [C] Students completed the background questionnaire, mathematics test and English test. [D] Survey students present at time of Wave 2 (and who were present in Wave 1) measured by completion of at least one test instrument at time of Wave 2.

The distribution of survey students across Grade 7 and Grade 8 is shown in Table 2.2. Students are distributed fairly evenly between the two grades in all regions. In Addis Ababa, Afar and Somali sites, there are slightly more students in Grade 8, counter to the trend in all other regions. The composition of the sample by grade did not change much between Waves 1 and 2.

Table 2.2. *Grade distribution of survey students, by location and survey wave*

Location	Survey students at Wave 1					Survey students at Wave 2				
	Grade 7		Grade 8		Total	Grade 7		Grade 8		Total
	N	%	N	%	N	N	%	N	%	N
Addis Ababa sites	1,119	49	1,183	51	2,302	1,031	50	1,026	50	2,057
Amhara sites	608	51	577	49	1,185	535	52	494	48	1,029
Oromia sites	579	56	462	44	1,041	519	54	435	46	954
SNNP sites	1,267	53	1,139	47	2,406	1,137	53	991	47	2,128
Tigray sites	908	53	799	47	1,707	842	53	733	47	1,575
Somali sites	1,251	49	1,320	51	2,571	1,048	48	1,133	52	2,181
Afar sites	478	49	492	51	970	395	49	416	51	811
Total	**6,210**	**51**	**5,972**	**49**	**12,182**	**5,507**	**51**	**5,228**	**49**	**10,735**

2.2 Student characteristics

2.2.1 Gender

The sample includes almost exactly the same number of female students as male students (5,958 females and 5,907 males). When disaggregated by region, moderate to large differences exist between the numbers of males and females (Table 2.3). In Somali and Afar this favours males (16 percentage point and 5 percentage point differences, respectively) while in all other regions the difference favours females, with gaps of between 3 and 10 percentage points. Gaps in terms of a Gender Parity Index (GPI) highlight the regional differences.[18]

Table 2.3. *Students' gender, by location*

Location	Survey students at Wave 1							
	Female		Male		No response		Total	GPI *
	N	%	N	%	N	%	N	
Addis Ababa sites	1,261	55	1,029	45	12	1	2,302	1.23
Amhara sites	633	53	540	46	12	1	1,185	1.17
Oromia sites	516	50	468	45	57	5	1,041	1.10
SNNP sites	1,180	49	1,081	45	145	6	2,406	1.09
Tigray sites	879	51	819	48	9	1	1,707	1.07
Somali sites	1,042	41	1,475	57	54	2	2,571	0.71
Afar sites	447	46	495	51	28	3	970	0.90
Total	**5,958**	**49**	**5,907**	**48**	**317**	**3**	**12,182**	**1.01**

Notes: * Gender Parity Index, based on attendance of survey sample.

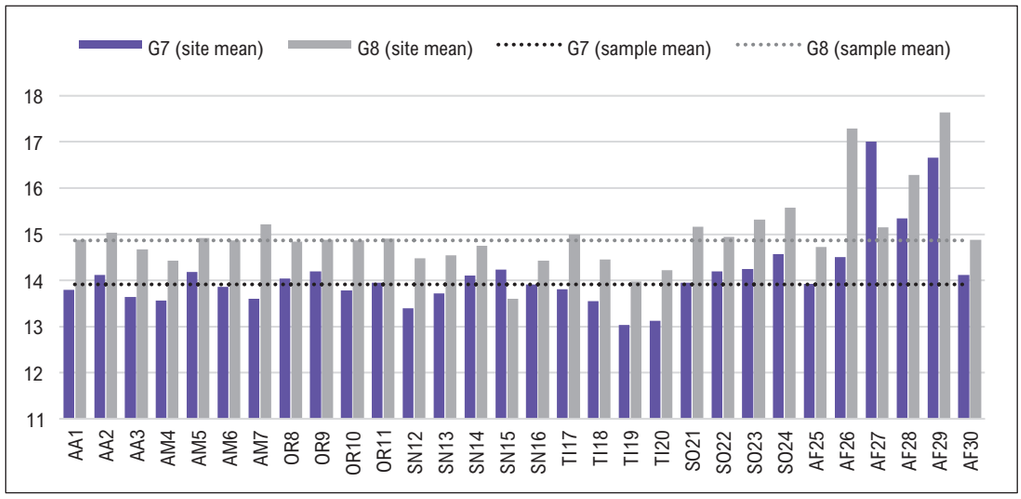
2.2.2 Age

Average student age across the sample is 14.4 years and this varies by grade, region and site (Figure 2.0). In the majority of sites, survey students are within one year of the correct age-for-grade, with the average age of the Grade 7 sample equal to 13.9 years and of the Grade 8 sample equal to 14.8 years (at the start of the school year, when student ages were recorded, age-for-grade would be 13 and 14 for Grade 7 and Grade 8, respectively).

In two sites, SNNP site 15 and Afar site 27, students in Grade 7 report a higher average age than their counterparts in Grade 8 but elsewhere students enrolled in Grade 8 are older. In general, students in the Somali and Afar sites are older than the sample average, while students in Tigray sites are younger than the sample average for each grade.

18 These GPI figures are calculated based only on student enrolments, assuming an equal number of males and females in each site. They are not calculated based on the age-specific population of females or males in each site, which is not known.

Figure 2.0. *Average student age reported at Wave 1, by site and grade*



2.2.3 *Students' mother tongues*

Ethiopia has 91 registered linguistic groups and students speak a variety of languages. Table 2.4 shows that three-quarters of survey students report one of three languages as the main language spoken at home: Amharic (47 per cent), Af-Somali (14 per cent) and Tigrigna (14 per cent).[19] There are important within-region variations in the main language spoken at home. For example, in the SNNP sites, 32 per cent of students speak a language from that region, but 54 per cent of students report speaking Amharic as the main language at home. A similar feature is observed in Afar sites.

Conversely, in Tigray and Amhara sites, much more homogeneity is observed, with 97 per cent and 98 per cent of students, respectively, speaking Tigrigna or Amharic as the main language in the home. These language variations between regional sites affect the links between mother tongues, languages of the home, languages of the teacher, and languages of instruction in the classroom.

19 Afaan Oromo is the Ethiopian language with the largest number of mother tongue speakers and so, for a nationally representative sample, would feature in a list like this. However, the sentinel site sampling approach used in our survey is not nationally representative. By the time of our survey, the number of students in schools within Oromia sites was small and neither were there many Afaan Oromo mother tongue speakers in our other sites.

Table 2.4. *Students' mother tongue, by location*

Location		Main language spoken at home (> 1%)								
		Afar Af	Amarigna	Afaan Oromo	SNNP languages [A]	Af-Somali	Tigrigna	Total of < 1 % [B]	No response [B]	Survey total
Addis Ababa sites	N	40	2,040	75	47		44	17	39	2,302
	%	2	89	3	2		2	< 1	2	100
Amhara sites	N	13	1,160					5	7	1,185
	%	1	98					< 1	< 1	100
Oromia sites	N	12	237	721	11			2	58	1,041
	%	1	23	69	1			< 1	6	100
SNNP sites	N	43	1,291	41	775			35	221	2,406
	%	2	54	2	32			1	9	100
Tigray sites	N						1,653	25	29	1,707
	%						97	1	2	100
Somali sites	N		486	32	69	1,715		50	219	2,571
	%		19	1	3	67		2	9	100
Afar sites	N	439	475	11				11	34	970
	%	45	49	1				1	4	100
Total	**N**	**570**	**5,691**	**881**	**921**	**1,727**	**1,712**	**73**	**607**	**12,182**
	%	**5**	**47**	**7**	**8**	**14**	**14**	**<1**	**5**	**100**

Notes: [A] The following languages have been grouped together under the category 'SNNP languages' for ease of presentation: Guragigna, Hadiyyisa, Sidaamu Afoo, Silt'e and Wolaytta. [B] In this column, all data are reported for completeness.

The majority of students report being able to have a conversation in at least one language other than their mother tongue. Perhaps unsurprisingly, rates are highest in the regional sites with the greatest diversity of mother tongues, as shown in Table 2.5. In five regions, the rate stands at around two-thirds, but in Tigray and Amhara, it falls to 26 per cent and 17 per cent, respectively.

Table 2.5. *Students who report speaking a language in addition to their mother tongue, by location*

Location	Students speak only home language		Student reports being able to have a conversation in at least one additional language		Missing data [A]		Total
	N	%	N	%	N	%	N
Addis Ababa sites	887	39	1,401	61	14	1	2,302
Amhara sites	979	83	199	17	7	1	1,185
Oromia sites	276	27	717	69	48	5	1,041
SNNP sites	563	23	1,692	70	151	6	2,406
Tigray sites	1,251	73	446	26	10	1	1,707
Somali sites	830	32	1,681	65	60	2	2,571
Afar sites	305	31	643	66	22	2	970
Total	**5,091**	**42**	**6,779**	**56**	**312**	**3**	**12,182**

Notes: [A] Data are 'missing' (a) if the student did not complete the questionnaire, or (b) if the student did not report a home language and reported only one total language in which he/she can have a conversation, or (c) if the student did not complete the question on languages that he/she can speak even if they reported a home language.

2.2.4 *Other characteristics*

The student questionnaire contained questions on students' households' durable assets for the purposes of measuring economic status. A composite score computed from these indicators is employed as a proxy measure for overall household economic advantage. Table 2.6 shows the percentage of students who fall into the most disadvantaged quintile in each region. Only 3 per cent of students in Addis Ababa sites fall into the most disadvantaged group, a proportion which rises to 44 per cent in Tigray sites.[20]

Students were asked about their general health and well-being and around one-third report one or more regular sight problem, hearing problem, headache, fever or stomach problem. In total, 16 per cent of students had lost one or both parents (were single or double orphans), and this was more common in Addis Ababa and Afar, at 20 per cent in each region.

We asked students whether they slept in the same house or compound all year round, whether they moved locations with animals, or moved locations for other reasons. We interpret movement with animals as a possible indicator of students with a pastoralist lifestyle and found the greatest numbers of these students in the SNNP and Afar sites.

Almost two-thirds of the students reported spending time on a 'usual school day' working in the family business or farm and three-quarters doing chores or caring for family members. Working for pay on a school day was found to be less common, but still 33 per cent of students reported working for pay on a 'usual school day'. Rates for all types of 'work' are lowest in Addis Ababa. 'Work' appears to be most common in Tigray sites, a finding no doubt related in part to household poverty. There are also gendered dimensions to time spent on different work activities which will be explored in future research.

Table 2.6. *Selected indicators of student backgrounds, by location*

Location	Students in 'poorest' quintile (%) [A]	Students reporting at least one health problem (%) [B]	Students reporting moving with animals (%) [C]	Students who report having lost at least one parent (%)	Students who report having moved to live with other people so closer to school (%)	On a usual school day, student reports spending some time on the following:[D]		
						Working in family business or on farm (%)	Doing chores or caring for family (%)	Working for pay or a wage (%)
Addis Ababa sites	3	38	1	20	10	30	51	11
Amhara sites	31	47	1	14	6	86	87	49
Oromia sites	19	27	6	17	11	89	86	36
SNNP sites	16	34	8	18	18	66	76	40
Tigray sites	44	25	6	10	4	87	87	63
Somali sites	18	43	2	15	8	47	70	21
Afar sites	22	33	9	20	7	62	63	25
Total	**20**	**36**	**4**	**16**	**10**	**62**	**73**	**33**

Notes: [A] This score is created by averages of Principal Components Analysis (PCA) of student-reported household durable assets, including telephone, radio, television, bicycle, car/truck, table, chair, fridge, bed with mattress. [B] Student reports one or more of the following health problems: sight problems, hearing problems, headaches, fever, stomach problems. [C] Student reports not sleeping in the same household or compound all year round, but rather moving with their animals. [D] Students were asked to detail how much time they spent on different work-related activities. Here, we have grouped all students who report spending anywhere between 'less than one hour' and 'more than four hours' into one category, omitting only those who report 'none'.

20 Young Lives sites are not randomly selected and our sites in Tigray are particularly poor relative to the rest of the sample and relative to a Tigray average. This must be borne in mind when interpreting results. The 30 school survey sites were selected purposively to ensure that the survey reflected the cultural and geographic diversity of the country, including urban–rural differences, but with a pro-poor bias and a focus on areas with food insecurity (see Outes-Leon and Sanchez 2008 for further details).

Students also reported on their educational aspirations and on the educational status of their parents (Table 2.7). Not all students are likely to know their parents' educational attainment, so two types of question were asked and can serve as indicators: one about parents' ability to read and write (in any language) and one about parents' educational attainment. Across the sample, students report that 63 per cent of mothers and 72 per cent of fathers can read or write, and similar shares have completed basic education.

Attainment of basic education is far higher in urban sites than it is in rural sites and is higher among fathers than mothers. The lowest rate of basic education attainment is among mothers in Amhara (25 per cent) and the highest is among fathers in Addis Ababa (89 per cent). In all sites, students report very high aspirations for future educational attainment: 80 per cent expect to complete university, with only moderate variation across sites (see also Tafere 2014).

Table 2.7. *Selected indicators of students' household education, by location*

Location	Mother can read and write (%)	Mother completed basic education (%) [A]	Father can read and write (%)	Father completed basic education (%) [A]	Students that expect to complete university (%)
Addis Ababa sites	72	78	75	89	85
Amhara sites	38	25	64	43	78
Oromia sites	63	61	70	76	78
SNNP sites	69	73	77	85	75
Tigray sites	50	36	74	49	66
Somali sites	54	58	74	80	87
Afar sites	43	47	58	65	87
Rural sites	58	32	64	45	68
Urban sites	43	67	75	83	84
Total	**63**	**58**	**72**	**73**	**80**

Notes: [A] Wherever a student reports mother or father education beyond Grade 4 (e.g. more than 'up to Grade 4'), counted as basic education.

2.3 Students' educational status and histories

The average age of enrolment in school overall is 6.9 years, against a policy expectation of 7 years, and ranges from 6.6 in SNNP sites to 7.2 in Somali sites (Table 2.8). Sixty per cent of students report having attended pre-primary education. This is considerably higher than would be expected by a random sample of Grade 7 or Grade 8 students in Ethiopia. Nationally, gross enrolment rates in pre-primary education were less than 5 per cent when this cohort passed that level (Ministry of Education 2009).[21]

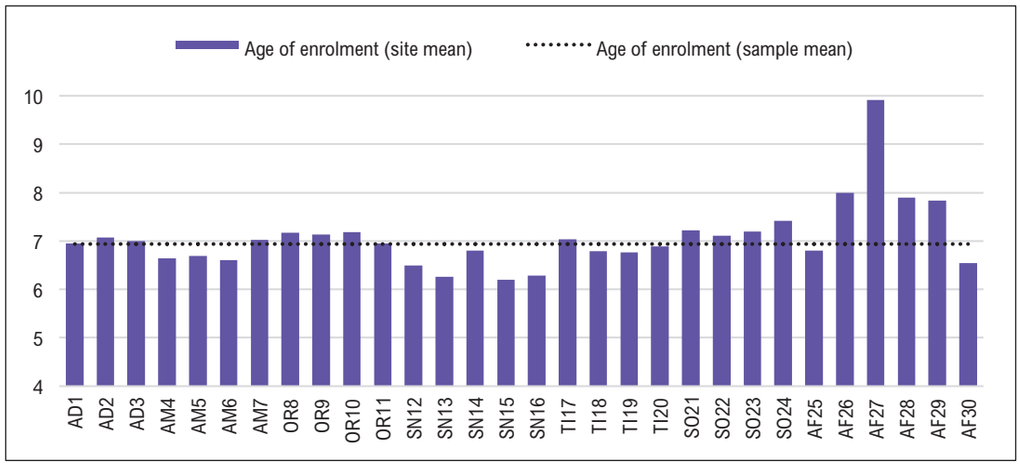
21 Pre-primary education could include formal and non-formal religious education, government-provided kindergarten or other programmes, and kindergarten provided by private or independent organisations.

Table 2.8. *Selected indicators of students' educational history and status, by location*

Location	Mean age of student enrolment (years)	Students enrolled by age 7 (%)	Students attended pre-primary education (%)	Students have attended same school since Grade 1 (%)	Student reporting ever having repeated a grade (%)	Student reporting ever having dropped out from a grade (%)
Addis Ababa sites	7.0	77	80	55	27	13
Amhara sites	6.7	82	25	69	21	19
Oromia sites	7.1	76	46	68	29	29
SNNP sites	6.6	81	71	45	31	23
Tigray sites	6.8	92	44	64	7	10
Somali sites	7.2	60	69	56	17	14
Afar sites	6.9	78	49	69	37	18
Rural sites	6.9	80	39	71	19	21
Urban sites	7.0	75	67	54	25	16
Total	**6.9**	**77**	**60**	**58**	**23**	**17**

Figure 2.1 presents site-level average age on school entry. In four of six Afar sites, students were closer to age 8 when they started school. In all sites except Afar site 27,[22] average age of enrolment is within a year of the policy expectation.

Figure 2.1. *Average age when starting school, by site*



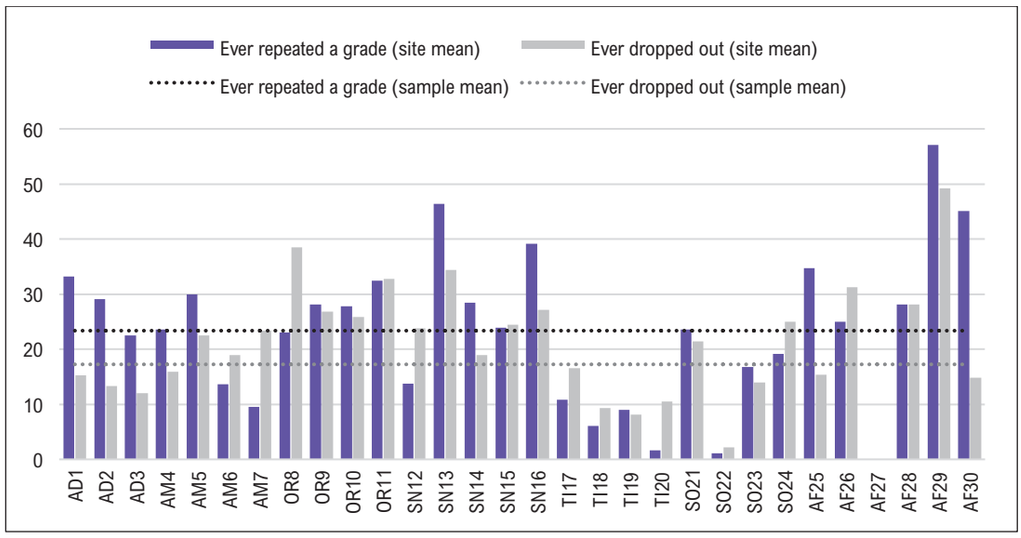
Across the sample, 23 per cent of students report having repeated a grade and 17 per cent of students report ever having dropped out.[23] There is, however, quite a lot of variation in grade repetition and dropout history by region and site. Figure 2.2 presents percentages of

22 Afar site 27, in Figure 2.1 and Figure 2.2 and elsewhere, is something of an outlier. It should be borne in mind when interpreting results that this site includes only one school in the survey sample.

23 In this survey 'dropout' means that a student terminated their learning within the school year, but it does not mean that they have permanently left education. For example, a student who decides after four months of term that they can no longer continue would 'dropout'. This individual could come back to school in the next academic year and depending on their school's policy might re-enrol in the grade that they left, or enrol in the grade that they would have reached had they not dropped out. This is a slightly different definition from that used in administrative datasets which compute dropout as a residual of progression and repetition.

students by site who have repeated a grade or have ever dropped out. Rates are particularly high in sites in Oromia, SNNP and Afar and particularly low in sites in Tigray (which is reflected in the lower average student age in these Tigray sites).

Figure 2.2. *Incidence of grade repetition and dropout, by site*



2.3.1 *Language of instruction in school*

Ethiopian primary schools teach predominantly in students' mother tongues in the earliest grades (mostly Grades 1 to 4), but in higher grades the number of languages of instruction decreases. Most students move away from learning in their mother tongue, especially for mathematics and science subjects, but this varies according to regional policy. In our sample, only five languages of instruction were used for mathematics in Grade 7 and Grade 8 and almost two-thirds of students (64 per cent) learn in English. Table 2.9 provides details of the language of instruction for survey students, by region. It should be noted that the language of instruction reported is not necessarily the main or only language used in the classroom.

Table 2.9. *Language of instruction for mathematics, by location*

Location	Language of instruction for mathematics (Grade 7 and Grade 8)				
	Afaan Oromo	Af Somali	Tigrigna	Amharic [A]	English
Addis Ababa sites	-	-	-	-	100
Amhara sites	-	-	-	-	100
Oromia sites	85	-	-	15	-
SNNP sites	-	-	-	-	100
Tigray sites	-	-	100	-	-
Somali sites	-	65	-	-	35
Afar sites	-	-	-	-	100
Total	**7**	**14**	**14**	**1**	**64**

Notes: [A] Two schools in Oromia region include 'Amharic' sections in which mathematics is taught in Amharic, using Amharic textbooks at Grade 7 and Grade 8.

Approximately 33 per cent of students learn in the same language that they speak at home. This varies slightly by grade, but substantially by region, as shown in Table 2.10. In Oromia, Tigray and Somali sites, most students learn mathematics in the language that they speak at home. The vast majority of students in survey schools in Afar, SNNP, Amhara and Addis Ababa are not learning mathematics in the language that they speak at home. In these four regions, mathematics instruction is officially in English, as a transition towards secondary grades – although note once again that this does not necessarily mean that English is the main or only language used in the classroom.

Table 2.10. *Mathematics instruction in mother tongue, by location*

Location	Percentage of students learning mathematics in the same language they speak at home		
	Grade 7	Grade 8	Total
Addis Ababa sites	6	1	3
Amhara sites	0	0	0
Oromia sites	79	85	82
SNNP sites	0	1	0
Tigray sites	97	97	97
Somali sites	51	61	56
Afar sites	0	0	0
Total	**33**	**33**	**33**

2.4 Student enrolment and attendance

The longitudinal nature of the survey enables an examination of changing patterns of enrolment and attendance over the course of a single school year. Table 2.1 presented the total number of students enrolled in survey sections at the time of Wave 1 and the numbers present at enumeration, constituting the student-level survey sample. These survey students were then followed up at Wave 2, which collected information on their enrolment status, whether they were present at the time of the survey, and their number of days of absenteeism over the course of the school year.

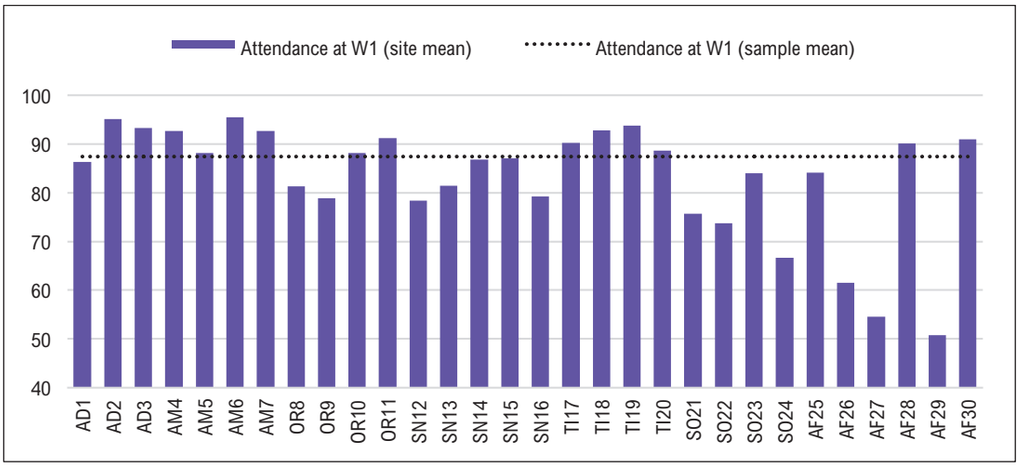
Overall, 87 per cent of students enrolled were present at the time of the Wave 1 survey (Table 2.11). Disaggregated by region, this figure ranges from 81 per cent in Afar sites to 94 per cent in Addis Ababa sites. Within-region variation in the attendance of students is also found (see Figure 2.3). Some Somali and Afar sites have particularly low rates of attendance, at around 50 per cent in Afar site 29.[24]

24 Low rates of attendance at time of the survey may also reflect less accurate school registers in the first few weeks of the school year.

Table 2.11. *Attendance at Wave 1 of students enrolled on the school register, by location*

Location	Children enrolled on register who were present in school at time of Wave 1				
	Present		Absent		Total
	N	%	N	%	N
Addis Ababa sites	2,302	94	156	6	2,458
Amhara sites	1,185	92	103	8	1,288
Oromia sites	1,041	87	153	13	1,194
SNNP sites	2,406	85	432	15	2,838
Tigray sites	1,707	92	150	8	1,857
Somali sites	2,571	83	543	17	3,114
Afar sites	970	81	224	19	1,194
Total	**12,182**	**87**	**1,761**	**13**	**13,943**

Figure 2.3. *Attendance at Wave 1 of students recorded on the register, by site*



Looking at student attendance at both waves, 88 per cent of students present at Wave 1 were also present at Wave 2 (Table 2.12). This ranges from 84 per cent in Afar sites to 92 per cent in Tigray and Oromia sites.

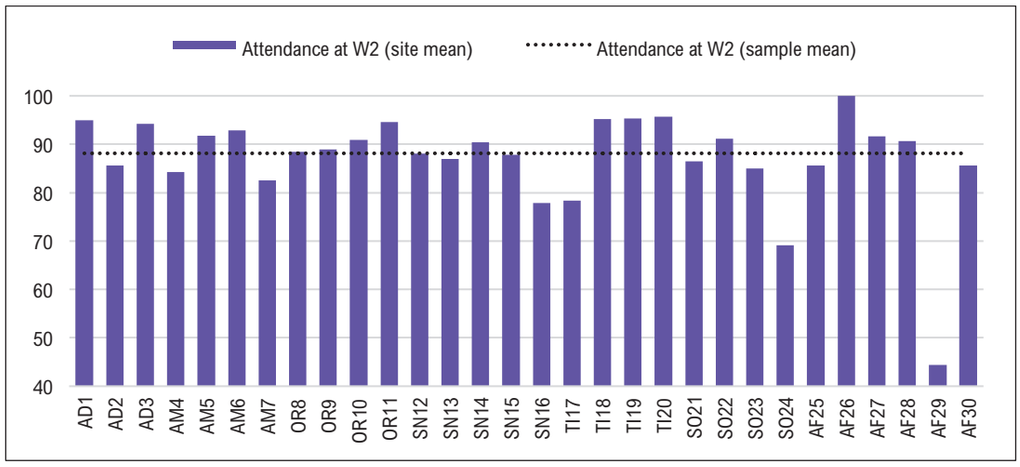
Table 2.12. *Student attendance at Wave 2, by location*

Location	Survey students present at Wave 1 attendance at time of Wave 2				
	Present at Wave 2 [A]		Not present at Wave 2		Total
	N	%	N	%	N
Addis Ababa sites	2,057	89	245	11	2,302
Amhara sites	1,029	87	156	13	1,185
Oromia sites	954	92	87	8	1,041
SNNP sites	2,128	88	278	12	2,406
Tigray sites	1,575	92	132	8	1,707
Somali sites	2,181	85	390	15	2,571
Afar sites	811	84	159	16	970
Total	**10,735**	**88**	**1,447**	**12**	**12,182**

Notes: [A] Survey student completed at least one of the Wave 2 survey instruments.

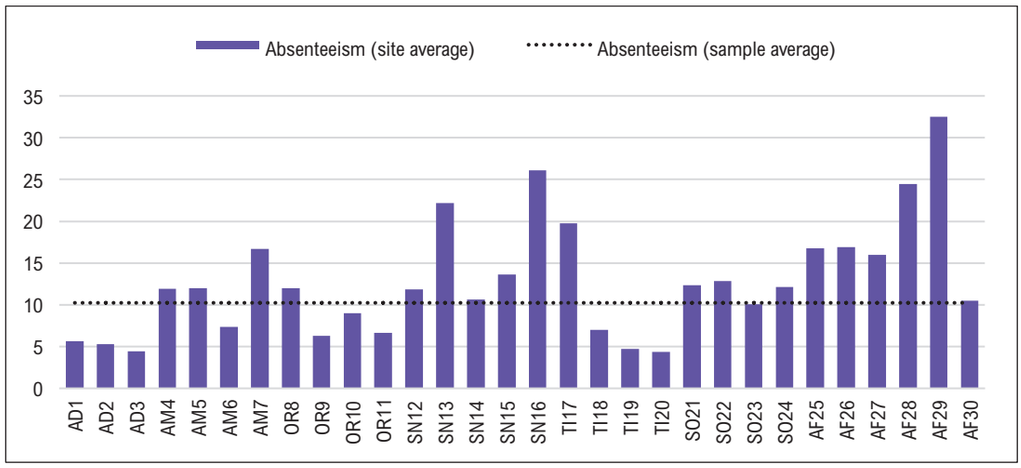
Disaggregation by site in Figure 2.4 shows substantial site-level variation in the share of survey students present at Wave 2. In Oromia sites, attendance at Wave 2 is consistently higher than the sample average. All students remain present in Afar site 26, compared to below 50 per cent in Afar site 29.

Figure 2.4. *Attendance of survey students at Wave 2, by site*



Data on student absenteeism during the academic year were collected from school administrative records. Overall, students were absent for an average of 10 per cent of the time period measured, ranging from 5 per cent of total time in the Addis Ababa sites to 17 per cent in the Afar sites.[25] Significant site-level variation is found (Figure 2.5), with students in four sites being absent for more than 20 per cent of the time period (i.e. one day per week).

Figure 2.5. *Absenteeism by survey site*



25 Note that administrative records at schools are generally kept by the 'home teacher' for each section. There is no standardised approach to recording attendance and so efforts were made during data collection to allow for flexibility in teacher reporting. In addition, some home teachers will keep records from both semesters, others will keep records for only one semester. These estimates exclude students for whom very little attendance data (fewer than five weeks) are recorded.

2.5 School, director, section and teacher characteristics

2.5.1 School and school director characteristics

Table 2.13 presents selected school and director characteristics by region. Survey schools range in age but are generally well established, with the average age for each region at more than 20 years. The typical director in survey schools is a 35-year-old male with a university degree and approaching ten years of experience as a school director. Only in Addis Ababa and SNNP sites are there survey schools with female directors.

Table 2.13. *Selected school and school director characteristics, by location*

Location	Mean age of school (years)	Mean age of director (years)	Male director (%)	Director has university degree (%)	Mean years as a director
Addis Ababa sites	36	39	67	89	10
Amhara sites	26	37	100	43	10
Oromia sites	27	33	100	83	6
SNNP sites	30	31	92	92	6
Tigray sites	34	42	100	86	12
Somali sites	22	38	100	57	11
Afar sites	22	27	100	50	6
Total	**28**	**35**	**94**	**71**	**9**

Note: This table provides statistics at the school level (n=63).

The availability of school facilities and resources varies considerably by region (see Table 2.14). Most schools have a pedagogical centre (a room in which teaching and learning materials are prepared, stored and shared), electricity on the day of the survey, a sports or play area, and a library. Most schools do not have a laboratory, a teachers' lounge or computers for students' use. Schools in Afar are less well equipped than the average school in every resource/facility listed.

Table 2.14. *Selected school facilities and resources, by location*

Location	Pedagogical centre (%)	Electricity on day of survey (%)	Sport/play area (%)	Laboratory (%)	Library (%)	Teachers' lounge (%)	Computers for students' use (%)
Addis Ababa sites	100	78	56	100	100	67	78
Amhara sites	57	57	100	43	57	14	29
Oromia sites	83	50	100	33	50	0	17
SNNP sites	75	75	92	58	75	67	17
Tigray sites	100	86	71	86	100	14	43
Somali sites	79	64	57	0	21	29	29
Afar sites	38	50	75	13	38	13	25
Total	**76**	**67**	**76**	**44**	**60**	**33**	**33**

Note: This table provides statistics at the school level (n=63).

The Ministry of Education and Regional Education Bureaus classify certain schools as 'cluster resource centres' and require these schools to support their neighbours.[26] A cluster resource centre will typically support five satellite schools, and will usually have more facilities and resources and a more experienced workforce. It will also usually offer a full cycle (i.e. eight years) of primary schooling and serve as a feeder into secondary schools.

In addition, since 2014, the Ethiopian education administration has been rolling out a school inspection service which is now capable of conducting regular inspections of service delivery. Schools will self-inspect periodically following standardised guidelines and an independent inspectorate will visit each school every one to two years.

Ratings from these processes (resource centre status and/or inspection rating) can be used as indicators of school quality. Just over half of the survey schools (54 per cent) are cluster resource centres (Table 2.15). In a random sample of primary schools we would expect one in six to be a cluster resource centre, but this higher rate might reflect the fact that the survey sampled only schools with Grade 7 and/or Grade 8 sections and it is these 'full-cycle' schools that are often better-established and selected as cluster resource centres.

Across the sample, schools' self-inspection and independent-inspection ratings (which sit on a scale from 1, worst to 4, best) are similar, with independent inspection ratings varying from an average of 2.0 in SNNP sites to 3.1 in Addis Ababa sites.

Some schools in Addis Abada, SNNP, Somali and Afar offer full-day teaching at Grade 7 and/or Grade 8. In all survey schools from Amhara, Oromia and Tigray sites, teaching takes place during either a morning or an afternoon shift, reducing the contact hours between schools and students.

Table 2.15. *Selected school quality indicators, by location*

Location	School is a cluster resource centre (%) [A]	Mean self-inspection rating [B]	Mean independent-inspection rating [C]	Schools that teach full day at Grade 7 or 8 (%)
Addis Ababa sites	44	2.8	3.1	89
Amhara sites	43	3.0	2.8	0
Oromia sites	33	3.0	2.6	0
SNNP sites	42	1.7	2.0	50
Tigray sites	43	3.0	2.6	0
Somali sites	71	2.0	2.2	36
Afar sites	88	2.4	2.7	13
Total	**54**	**2.4**	**2.5**	**32**

Notes: This table provides statistics at the school level (n=63). [A] Primary schools are grouped into clusters of approximately six. The resource centre is usually a more established and better equipped school which supports its satellites. [B] For all schools that have self-inspected. [C] For all schools that have been inspected.

Water and sanitation conditions in survey schools are often good in terms of nominal availability. Almost all schools have pit latrines or flush toilets for students and 60 per cent have an improved water source at the school (Table 2.16). On the other hand, about one-quarter of schools have no access to water whatsoever, only 56 per cent of survey schools had access to water on the day of the survey, and less than one-third had a place for

26 For an overview of school clusters and resource centres, see: http://unesdoc.unesco.org/images/0015/001597/159776e.pdf.

students to wash their hands – with this an acute problem in Tigray sites. Toilets are generally separated by gender (Table 2.17), although this is not always the case. Females have a place to wash menstrual rags or a private space to wash in only a handful of schools (ten and nine schools, respectively).

Table 2.16. *Selected school water and sanitation facilities, by location*

Location	Pit latrines or flush toilets (%)	Any water source at school (%)	Improved water source (%) [A]	Water available at time of survey (% of those with water)	Place to wash hands (%)
Addis Ababa sites	100	100	100	78	89
Amhara sites	86	43	29	43	0
Oromia sites	100	83	83	67	0
SNNP sites	100	75	75	67	42
Tigray sites	100	57	57	0	0
Somali sites	71	79	29	57	36
Afar sites	100	88	63	63	0
Total	**92**	**76**	**60**	**56**	**29**

Notes: This table provides statistics at the school level (n=63). [A] Drawing on the WHO/UNICEF Joint Monitoring Programme for Water Supply and Sanitation categorisation of improved and unimproved drinking water, the following response categories are coded as 'improved' sources of drinking water: protected well or spring, tube well or borehole, piped water.

Table 2.17. *Selected gender-specific school water and sanitation facilities, by location*

Location	Separate toilets for girls/boys (%)	Place for girls to wash menstrual rags (%)	Private space for girls to wash (%)
Addis Ababa sites	100	11	0
Amhara sites	86	14	14
Oromia sites	83	0	0
SNNP sites	100	33	42
Tigray sites	86	29	0
Somali sites	86	14	21
Afar sites	75	0	0
Total	**89**	**16**	**14**

Note: This table provides statistics at the school level (n=63).

2.5.2 Section and teacher characteristics

Table 2.18 presents selected section-level characteristics. Across the survey sample the average section contains 51 students and ranges from 44 in Amhara to 62 in Somali. Note that these numbers are based on official enrolment (from school registers) and should be interpreted along with the indications of attendance in Table 2.11 and Figure 2.3.

Indicators of instruction include average hours of mathematics teaching per week, the equivalent for English and the share of all sections for which all students have their own textbook. Teaching hours for mathematics and for English vary little apart from in Addis Ababa sites, in which survey students receive approximately one additional hour (about 30%

more contact time) per week.[27] Students in these sites also always have their own textbook, which is not the case everywhere. In Afar sites only 38 per cent of sections have students all with their own English textbook, and that rate is only slightly higher in Somali sites.

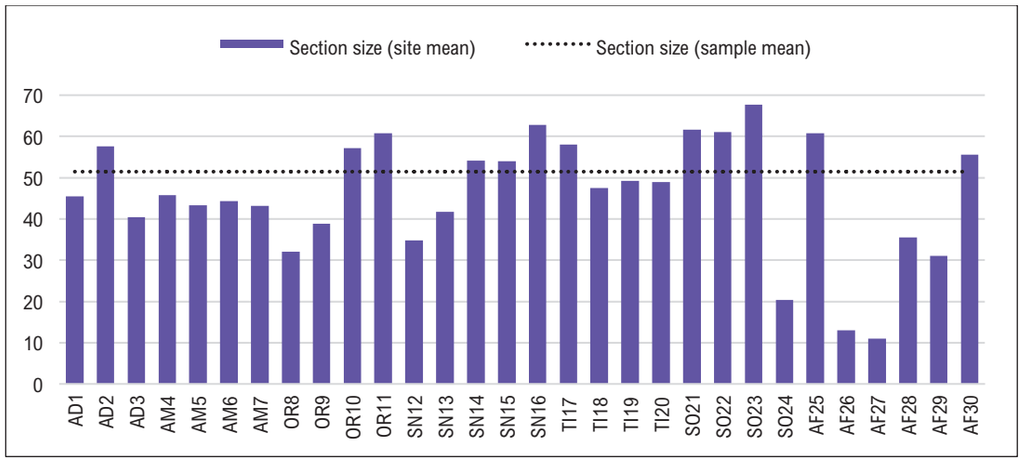
Table 2.18. *Selected section teaching and learning characteristics, by location*

Location	Mean class size (number of students) [A]	Mean hours of maths teaching per week (hours)	Mean hours of English teaching per week (hours)	All students in class have their own maths textbook (%)	All students in class have their own English textbook (%)
Addis Ababa sites	49	4.4	4.4	100	100
Amhara sites	44	3.3	3.3	86	93
Oromia sites	52	3.3	3.3	74	74
SNNP sites	51	3.4	3.4	77	68
Tigray sites	50	3.3	3.3	78	68
Somali sites	62	3.3	3.2	52	44
Afar sites	46	3.2	3.4	54	38
Total	**51**	**3.5**	**3.5**	**75**	**70**

Notes: This table provides statistics at the class level (n=271). [A] Based on enrolment, not attendance, at time of survey.

Figure 2.6 reports the average section size by site. Variation in section size between sites within a region is particularly pronounced in Afar and Oromia, ranging from 30 or fewer students per section up to over 50 students per section. In Somali, all except one site has a higher than sample-average section size, with one site approaching 70 students per section; and in SNNP three of the five sites are above average. Again, these numbers are based on official enrolment (from school registers) and should be interpreted along with the indications of attendance in Table 2.11 and Figure 2.3.

Figure 2.6. *Average section size based on enrolment, by site*



27 Within sites there are some schools that offer full-day instruction and more hours of instruction, but these are so few in number that they are lost in regional averages.

The survey sample included all teachers of mathematics and all teachers of English to students in Grade 7 and Grade 8 in survey schools (the subjects in which students completed repeated cognitive tests). In total, 211 teachers were interviewed. Tables 2.19 and 2.20 present selected indicators of teachers' background, training and experience, by subject.

The average age of survey English teachers (n=105) is 35 years, varying from 30 in Addis Ababa sites to 38 in Somali and Oromia sites. The vast majority of English teachers (86 per cent) specialised in English during their training, but this falls to 38 per cent in Amhara sites. Three-quarters of English teachers have completed post-secondary education (e.g. a three-year teaching diploma) or higher education. Of all English teachers, 69 per cent are male. Years of teaching experience range from eight to 19 years between regions, with a sample average of 14 years.

Table 2.19. *Selected English teacher characteristics, by location*

Location	English teachers				
	Mean teacher age (years)	Male teacher (%)	Mean years teaching experience (years)	Completed post-secondary or higher education (%)	Has specialised in English during training (%)
Addis Ababa sites	30	58	8	84	96
Amhara sites	37	55	15	97	38
Oromia sites	38	65	16	87	83
SNNP sites	37	75	17	96	77
Tigray sites	32	68	12	30	100
Somali sites	38	78	19	86	86
Afar sites	31	84	11	19	96
Total	**35**	**69**	**14**	**75**	**83**

Note: This table provides statistics at the class level (n=271).

Mathematics teachers in survey schools (n=106) are similar to English teachers in many ways. They are almost the same age (33 as opposed to 35 years old), most have completed post-secondary or higher education, have a similar number of years' teaching experience (12 compared to 14 years), and are specialists in the subject that they teach. The biggest difference between the two sets of teachers, on the indicators presented, relates to gender: of all mathematics teachers, 91 per cent are male, compared to 69 per cent of English teachers.

Table 2.20. *Selected mathematics teacher characteristics, by location*

Location	Maths teachers				
	Mean teacher age (years)	Male teacher (%)	Mean years teaching experience (years)	Completed post-secondary or higher education (%)	Has specialised in maths during training (%)
Addis Ababa sites	33	91	9	86	90
Amhara sites	29	86	10	52	28
Oromia sites	37	89	17	87	83
SNNP sites	35	93	14	95	89
Tigray sites	31	81	11	70	100
Somali sites	35	96	12	94	86
Afar sites	29	100	8	69	100
Total	**33**	**91**	**12**	**82**	**84**

Note: This table provides statistics at the class level (n=271).

3. Student cognitive achievement

This section presents student achievement in three cognitive domains: mathematics, English and Amharic. For all three, it includes an indicator of absolute achievement (levels) and for mathematics and English, thanks to the repeated measures design, it also indicates progress made within the academic year (gains).

3.1 Cross-sectional cognitive achievement: percentage correct

First, achievement is shown in terms of percentage correct on the Wave 1 mathematics and English tests and the Wave 2 Amharic test. Differences are presented between regions, between males and females, and between rural and urban sites. When interpreting results, note that the mathematics and English tests are norm-referenced, with a target average score of 50 per cent. This approach allows us to investigate differences between groups and differentiate between students at all levels of ability; we do not end up with lots of students with 0 scores or 100 per cent correct. Test scores do not, however, have any direct link to curriculum expectations, that is, we cannot set a threshold for proficiency at 20 per cent correct, 40 per cent correct, 60 per cent correct, or at any other level.

3.1.1 Mathematics

As the first indicator of achievement levels and differences between groups, across the survey sample, students scored an average of 41 per cent on the Wave 1 mathematics test (Table 3.0). Females scored 40 per cent, compared to males' 42 per cent, a difference equivalent to about one question on the 40-item test. In Amhara and Oromia sites, the gender difference stretches to 4 percentage points and in Somali sites it falls to 1 percentage point. Although there are roughly the same number of males and females in the survey sample, the importance of these achievement differences will depend on further investigation into the composition of each group (e.g. student home backgrounds, the schools they are attending and whether they reflect broadly equivalent draws from the population in each site, or not).

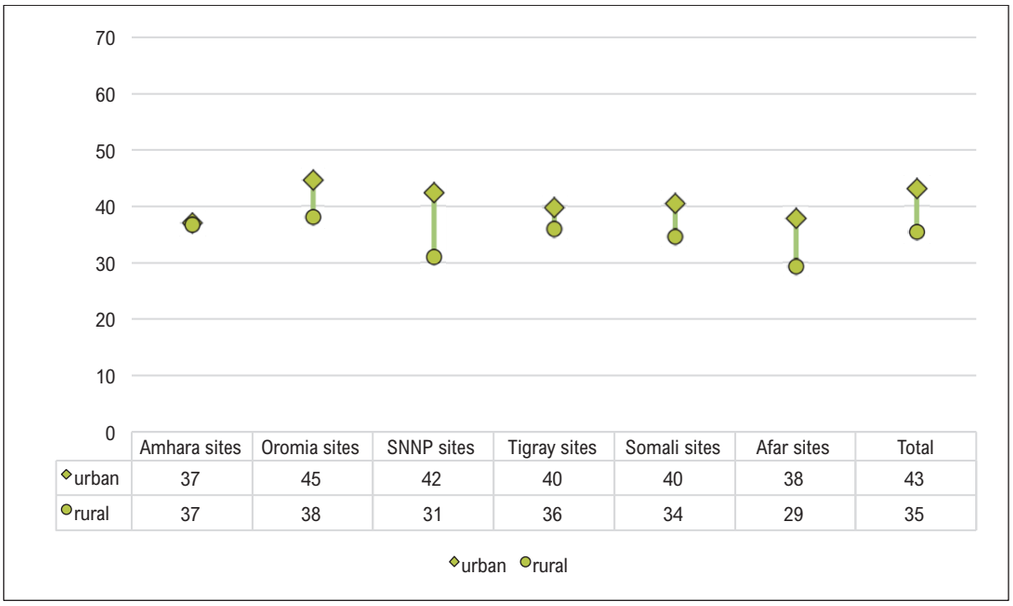
Table 3.0. *Student achievement on Wave 1 mathematics test, percentage correct, by location*

Location	All sites			Rural sites			Urban sites		
	F	M	T	F	M	T	F	M	T
Addis Ababa sites	49	51	50	-	-	-	49	51	50
Amhara sites	35	39	37	34	39	37	37	37	37
Oromia sites	40	44	41	37	40	38	42	49	45
SNNP sites	39	42	40	30	32	31	41	45	42
Tigray sites	36	39	37	34	38	36	40	40	40
Somali sites	39	40	40	32	36	34	40	41	40
Afar sites	36	39	37	29	30	29	36	39	38
Total	**40**	**42**	**41**	**34**	**37**	**35**	**42**	**44**	**43**

A far larger gap is shown between rural and urban sites in each region (Figure 3.0). Holding Addis Ababa as an exception (as it has only urban sites), in all regions except Amhara we find large difference in achievement between rural and urban sites. At the average this difference stands at 8 percentage points, rising to 11 percentage points in SNNP sites. In those SNNP sites, the average student in an urban area scores 35 per cent more in the mathematics test than his or her counterpart in a rural area (i.e. 11 percentage points divided by 31 per cent average in rural sites). When you layer rural and urban differences onto gender differences the largest gap – 15 percentage points – is between male students in urban SNNP, and female students in rural SNNP.[28] These students are learning within the same regional education system, yet their outcomes at age 14 or 15 are quite different.

Accepting that all gaps must later be interpreted in terms of student background, schools attended and so on, it is nonetheless important to note that there are large achievement differences between student groups.

Figure 3.0. *Achievement differences between rural and urban sites, Wave 1 mathematics test, by location*



	Amhara sites	Oromia sites	SNNP sites	Tigray sites	Somali sites	Afar sites	Total
urban	37	45	42	40	40	38	43
rural	37	38	31	36	34	29	35

3.1.2 English

Across the sample, students' average score was 48 per cent in the Wave 1 English test.[29] Table 3.1 breaks scores down by location and gender. Across all sites, there is no discernible gender gap: both males and females achieve 48 per cent. In all locations except Oromia and Afar sites, performance is the same for each group. In Somali rural sites and Tigray urban sites, females score slightly higher than males, but again the difference is negligible.

28 Urban enrolment in SSNP is dominated by students from schools in Hawassa.

29 Percentage correct scores are test dependent and are influenced by the difficulty of items. Performance on the English test should not be compared with performance on the mathematics test. In addition, percentage correct scores bear no relation to stated curricular expectations.

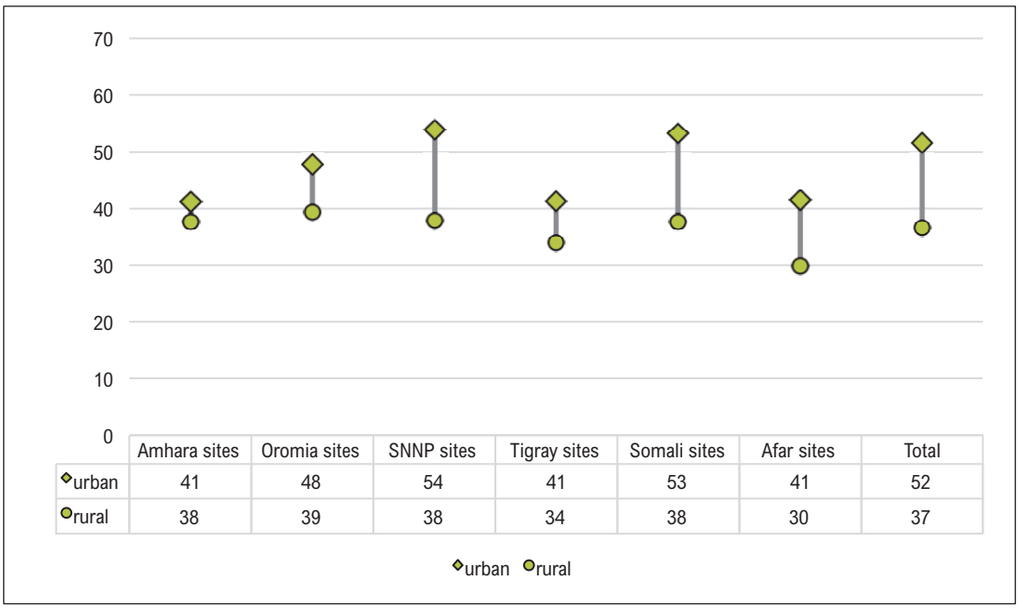
Table 3.1. *Student achievement on Wave 1 English test, percentage correct, by location*

Location	All sites			Rural sites			Urban sites		
	F	M	T	F	M	T	F	M	T
Addis Ababa sites	58	58	58	-	-	-	58	58	58
Amhara sites	39	39	39	38	38	38	41	41	41
Oromia sites	42	45	43	39	40	39	46	51	48
SNNP sites	50	50	50	37	39	38	54	54	54
Tigray sites	37	37	37	33	35	34	42	41	41
Somali sites	51	51	51	38	37	38	53	53	53
Afar sites	40	41	40	29	30	30	41	42	41
Total	**48**	**48**	**48**	**36**	**37**	**37**	**51**	**52**	**52**

The rural-urban gap identified in mathematics achievement exists – and is amplified – for English. Overall the difference stands at 15 percentage points, is highest in SNNP and Somali sites and is lowest, at 3 percentage points, in Amhara sites. Figure 3.1 highlights the often stark differences between rural and urban groups in each region.

Taking both mathematics and English achievement together – and recognising that all secondary subjects are taught in English – the chance that a male student from an urban site does well in further education is quite considerably higher than that for a female student learning in a rural site. Potential reasons for and wider implications of these differences will be a focus of further research.

Figure 3.1. *Achievement differences between rural and urban sites, Wave 1 English test, by location*



3.1.3 Amharic

The Amharic test was included to provide an indicator of students' Amharic proficiency; unlike the mathematics and English tests it was not norm-referenced. We expected many students – most likely those students with Amharic as a mother tongue – to score 100 per cent in the test. Across the sample the average score was 74 per cent, the median score 83 per cent and, as expected, 20 per cent of students achieved full marks.

Counter to results from mathematics and English, female students' average score of 76 per cent was higher than that for males, at 71 per cent (Table 3.2). Higher female performance is observed in both rural (smaller difference) and urban (larger difference) sites. Taking all students in rural sites, there is a 21 percentage point gap to their peers in urban sites. When interpreting these tables, note that there is not a consistent share of rural and urban sites per region or per language group. This will influence the gaps presented in Table 3.2.

Future Young Lives publications will break Amharic achievement down into proficiency bands (e.g. basic, proficient, advanced), and investigate the relationships between Amharic achievement, student background and performance in other cognitive tests. We will also investigate how Amharic proficiency varies by mother tongue and language of instruction model.

Table 3.2. *Student achievement on Wave 2 Amharic test, percentage correct, by location*

Location	F	M	T
Rural sites	61	60	60
Urban sites	84	79	81
Total	**76**	**71**	**74**

3.2 Repeated measures cognitive achievement: IRT

For mathematics and English, all items from Wave 1 and Wave 2 have been scaled concurrently using Item Response Theory (IRT) and then transformed onto a scale with average at 500 and standard deviation at 100 (see Box 3.0). The creation of a common scale between waves is made possible by a large number of common 'anchor' items between tests. The scaled scores offer no meaning in terms of the number of questions that a student answered correctly, but they do allow: (i) a 'truer' interpretation of differences in student achievement, because IRT – unlike a straightforward percentage correct score – does not assume that all items are equally difficult and worth one point; and (ii) a direct comparison of performance, on an interval scale, of achievement at Wave 1, at Wave 2, and in terms of the progress made during one academic year.

Box 3.0. *Scaling and reporting of results*

Results from tests in mathematics, English and Amharic are presented, depending on the purpose, either as 'raw' scores, percentage correct scores or 'interval scaled scores'. Interval scaled scores are computed using methods based on Item Response Theory (IRT) and are intended to provide a more precise measure of the underlying skill domains which are being assessed. The approach conceptualises the skill domains as 'latent traits' (e.g. mathematics proficiency), which are observed indirectly through response patterns to a set of question items.

IRT models are employed to simultaneously estimate both 'item' and 'person' parameters, where the first describe the characteristics of the test items – including 'item difficulty' – and the second describes the 'ability' of test-takers. 'Item difficulty' describes the probability that the item is answered correctly based on the responses among the sample, while 'person ability' depends on how many and which questions a student answers correctly. Using IRT modelling, we generate interval scaled scores which reflect not only the number of questions answered correctly by a student, but also the difficulty levels of the questions answered correctly. Interval scaled scores may be considered directly comparable estimates of students' levels of the relevant underlying skill (latent trait).

The 'person ability' parameter (latent trait) estimate is transformed onto a scale with average fixed at 500 and standard deviation fixed at 100. This convention, followed by international assessment studies including PISA and TIMSS, renders the scaled scores readily comparable in standard deviation terms.[30] For example, in respect of two students scoring 500 and 575, it can be immediately appreciated that the difference in scores is 0.75 standard deviations (since one standard deviation is defined as 100 points on the test-score scale). More detail on the technical procedures is provided in Part 4.

3.2.1 Mathematics

When looking at learning progress over the academic year, scores have been scaled together and so are directly comparable across time for individual students and for student groups. Wave 1 scores are presented for the full sample of students that were present at the beginning of the school year and Wave 2 scores are presented for all students from the initial sample that were present at the end of the school year (Figure 3.2). In order to calculate gain, both beginning and end of school year scores are required. As a result, when progress is reported, it includes only those students that were present at both survey waves. For this reason, 'progress' differs somewhat from the difference in averages between Wave 1 and Wave 2 when all students are included.[31]

Patterns of achievement at Wave 1 are consistent, albeit with minor differences, between this IRT-scaled score and the percentage correct presented in Table 3.0. The real value of the repeated measures design, however, is that it allows us to look at change over the school

30 The Programme for International Student Assessment (PISA) is a triennial international survey run by the OECD, which aims to evaluate education systems worldwide by testing the skills and knowledge of 15-year-old students. See: www.oecd.org/pisa. The Trends in International Mathematics and Science Study (TIMSS) is an international assessment of mathematics and science at the fourth and eighth grades that has been conducted every four years since 1995. See: http://timss2015.org.

31 This can reflect compositional effects (e.g. lower achievers at Wave 1 may be more likely to have dropped out).

year and begin to account for many of the student background and prior learning factors that affect achievement levels.

Across the survey sample, average gain over the course of the school year was 31 points, or just over a third of a standard deviation.[32] Achievement varied by location (Table 3.3), with highest performance at both waves (562 and 588 points, respectively) in Addis Ababa. At Wave 1, Amhara and Tigray sites were lowest performers but by Wave 2, Afar sites were lowest performers, reflecting slightly less progress in the year. Oromia sites made most progress, 43 points, over the academic year and that was from an already high starting point, thereby increasing the gap to most other sites.

Table 3.3. *Student achievement on Wave 1 and Wave 2 mathematics tests, IRT-scale, by location*

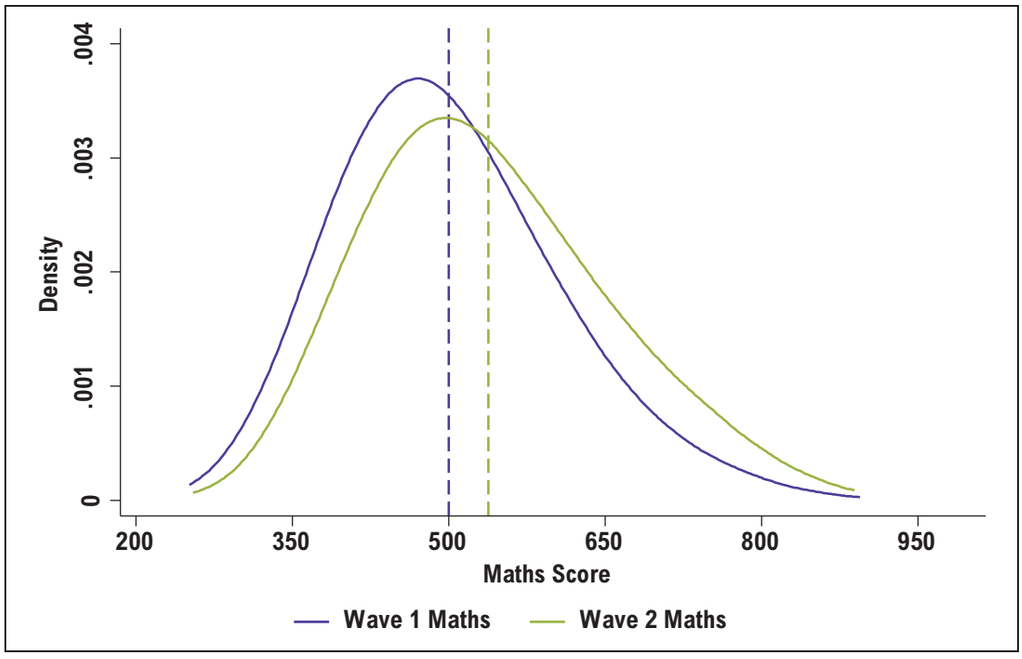
Location	Full sample at each wave		Only students present at both waves		
	Wave 1	Wave 2	Wave 1	Wave 2	Progress
Addis Ababa sites	562	588	565	588	22
Amhara sites	471	515	476	515	40
Oromia sites	497	543	500	543	43
SNNP sites	493	523	499	523	24
Tigray sites	471	517	475	517	42
Somali sites	495	535	503	535	33
Afar sites	473	505	477	505	28
Rural sites	459	492	462	492	30
Urban sites	514	553	522	553	32
Total	**500**	**538**	**506**	**538**	**31**

From a school-effectiveness perspective, even though they have lowest absolute performance, Afar sites may be doing reasonably well compared to those in Addis Ababa or SNNP. They have made a gain of 28 points within the academic year and schools in these areas are likely to be contending with greater challenges than in other sites. Similarly, all rural sites have made 30 points of progress over the academic year which, when compared with 22 points and 24 points of progress in Addis Ababa and SNNP sites, respectively, changes the perspective on achievement. Understanding what this progress means for mathematics proficiency and mastery is an issue for further research.

With these data on student progress it will be possible to investigate not only the factors that are associated with absolute levels of achievement, but also those factors which influence changes in learning over a school year. This type of factor may have stronger policy implications if it relates to aspects which the education system can affect.

32 In the progress section of Tables 3.3 and 3.4, the Wave 1 average score of students that were present for both tests is up to 506 and 504 points, respectively, reflecting non-random dropout of lower achievers between survey waves.

Figure 3.2. *Kernel density plots of student mathematics achievement (all students), by survey wave*



Note: * This kernel density plot includes all students that were present at Wave 1 and all students that were present at Wave 2. The purple vertical line reflects the Wave 1 average (500 points) and the green vertical line reflects the Wave 2 average (538 points).

3.2.2 English

In English, as with mathematics, patterns of achievement between regions at Wave 1 are consistent between this IRT-scaled score and the percentage correct presented in Table 3.1 (see also Figure 3.3). Regions retain their rank-order and gaps do not change substantially. The difference between rural and urban areas is 89 points, or nine-tenths of a standard deviation. This is a substantial 'urban dividend' that accrues mostly to sites in Addis Ababa, Somali and SNNP.

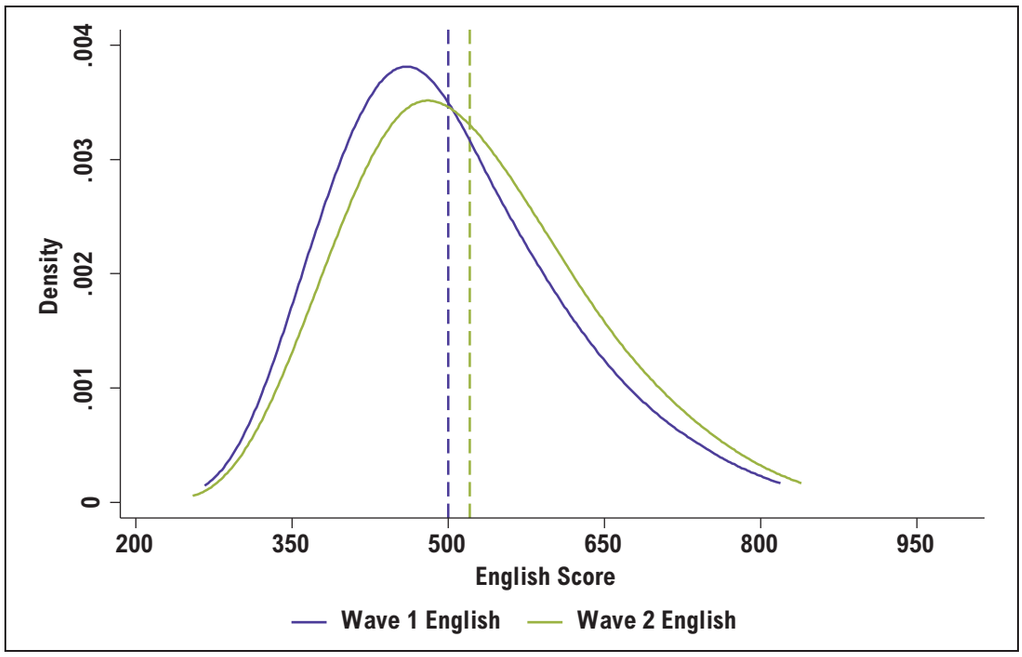
Student achievement in Amhara sites is low at Wave 1 (450 points), but when we look at progress, the change over the academic year is large at 31 points (Table 3.4). Somali sites also make significantly more progress than average over the academic year, and from a high starting point. Urban sites make more progress than rural sites: 20 points compared to 16 points. This tends to increase the already substantial gap between students based on location (up to 94 points by Wave 2) and increases the potential challenges around preparedness for secondary and further education.

Table 3.4. *Student achievement on Wave 1 and Wave 2 English tests, IRT-scale, by location*

Location	Full sample at each wave		Only students present at both waves		
	Wave 1	Wave 2	Wave 1	Wave 2	Progress
Addis Ababa sites	563	574	565	575	10
Amhara sites	450	481	451	482	31
Oromia sites	474	493	476	493	17
SNNP sites	518	535	523	538	15
Tigray sites	435	456	437	457	20
Somali sites	522	552	528	554	25
Afar sites	456	481	461	481	20
Rural sites	434	452	436	453	16
Urban sites	523	546	527	547	20
Total	**500**	**521**	**504**	**523**	**19**

Within the English test framework, items were classified by CEFR level, relating broadly to basic, independent and proficient users of the language. Future analysis will seek to understand student performance in terms of expected proficiency and will try to make sense of what 19 points of progress means in relation to this.

Figure 3.3. *Kernel density plots of student English achievement (all students), by survey wave*



Note: * This kernel density plot includes all students that were present at Wave 1 and all students that were present at Wave 2. The purple vertical line reflects the Wave 1 average (500 points) and the green vertical line reflects the Wave 2 average (521 points).

4. Technical information

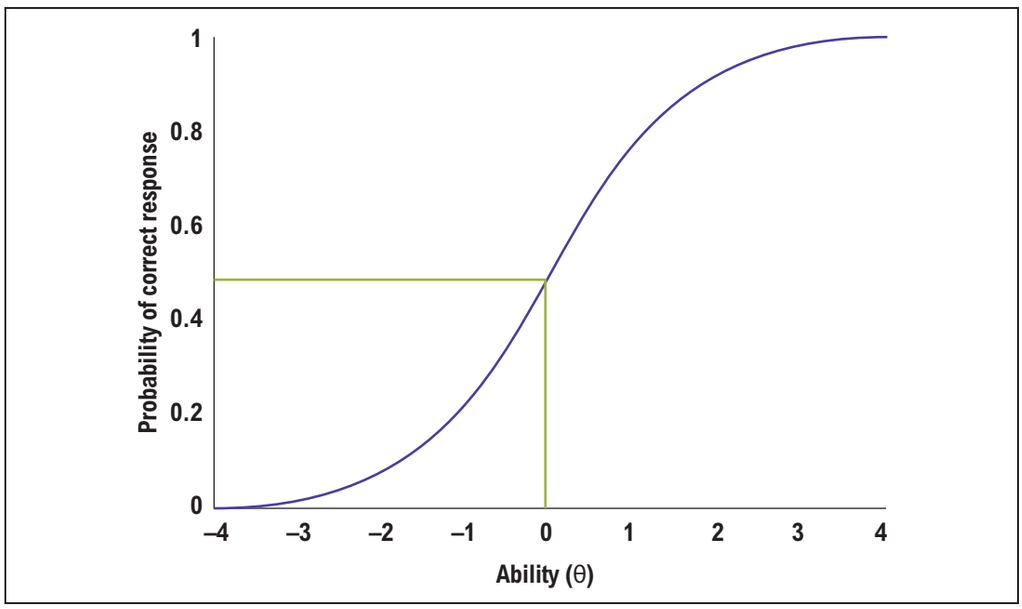
4.1 IRT scaling procedures (mathematics and English)

As described briefly in Section 3.2, data from tests in mathematics and English presented in this report have been scaled using psychometric procedures based on Item Response Theory (IRT). This is in order to provide interval-scaled measures of the relevant latent traits – the underlying skill proficiencies being examined by each test. We do not describe these procedures in detail here, but the reader is referred to Van der Linden et al. (2013) for a full discussion. Below we provide a non-technical summary of the approach taken to IRT scaling of the data.

In the case of the mathematics and English tests, each item follows a multiple choice format with a single correct answer, so that each item score is dichotomous, being either correct or incorrect. The latent trait modelling approach employed in all cases is the two-parameter IRT model. These models were implemented using the IRT suite of commands in STATA 14.

The two-parameter model simultaneously estimates two item parameters (difficulty and discrimination) and a single person parameter (ability). The difficulty parameter simply describes the probability that an item is answered correctly based on the response pattern among the sample of test-takers, while the discrimination parameter describes the rate at which the probability of answering an item correctly changes with student ability. Accordingly, the discrimination parameter is given by the slope of the item characteristic curve in Figure 4.0. An item characteristic curve represents the relationship between a latent trait (ability) and performance on an individual test item.

Figure 4.0. *Example item characteristics curve*



In this example, a student with the mean level of ability (a value of zero on the mean-centred scale below) has a 50 per cent probability of answering this item correctly. In turn, the difficulty parameter is zero (the item is at the mean level of difficulty). The ability axis scale measures both 'item difficulty' and 'student ability' in logits (log odds units), a log-transformed probability measure whereby one logit is the distance along the axis (increase in the value of the latent trait) which increases the odds of a correct answer by a factor of 2.718 (the base value of the natural logarithm).

In mathematics and English, two sets of tests were administered, with a sub-set of items being common to both tests (anchor or link items). As a result we are able to employ 'concurrent calibration' across items from both tests to estimate latent trait values for each student separately at the earlier and later time points during the school year, on a common interval scale (known as vertical scaling). This method includes all items administered across both tests and effectively includes each individual twice – first as a 'Wave 1' test-taker and second as a 'Wave 2' test-taker. This approach provides two ability estimates for each student, which allows calculation of 'progress made' between the two tests by simple subtraction of the earlier score from the later score.

As seen above, estimates of latent traits (proficiencies) from IRT models are outputted in the form of variables centred on zero reported in logits. Since the logit scale is an interval scale these values may be transformed onto any other interval scale while retaining the same properties. They are re-scaled for this report in order to aid interpretation, specifically to avoid negative proficiency scores and to ensure that all scores may be readily compared in standard deviation terms. In common with the approach of international studies such as PISA and TIMSS, we choose to rescale these variables to have a mean of 500 and a standard deviation of 100 for the Wave 1 data (Box 4.0). Data from Wave 2 tests are transformed specifically onto the Wave 1 scale. Accordingly, the Wave 2 scores can be readily interpreted by comparison with Wave 1 scores (in terms of progress).

Box 4.0. *Formulae to rescale student ability parameters to mean 500 and standard deviation 100*

T1: Student ability parameter (logit scale) at Wave 1

T2: Student ability parameter (logit scale) at Wave 2

M1: Mean of T1

SD1: Standard Deviation of T1

*Scaled Test Score Wave 1 = 500 + (100 / SD1) * (T1 - M1)*

*Scaled Test Score Wave 2 = 500 + (100 / SD1) * (T2 - M1)*

The design of multiple choice items included attention to the selection of 'distractor' options. These distractors serve as alternatives to the correct answer and are designed to reflect common misunderstandings of the concept in question. According to this approach, it is expected that students with weaker understanding will select distractor options somewhat more often than adopting an approach of random 'guessing'. For this reason – and after reviewing student selections of distractors – we chose not include a third parameter, known as the 'pseudo-guessing parameter' which is employed in some other studies (the three-parameter IRT model). This parameter is intended to account for the non-zero probability that

a very low proficiency student will select the correct answer to a multiple-choice question 'by guessing' and is defined as the lower asymptote of the item response function.

4.2 Pilot procedures

Test items and background questionnaires were pre-piloted and piloted between April – May 2016 (for Wave 1 instruments) and February – March 2017 (for Wave 2 instruments). Qualitative pre-piloting aimed to check the suitability of test items and questionnaire content, and to identify any issues with translation. Instruments were revised based on student and teacher feedback after the pre-pilot, and then a larger scale pilot was conducted.

The pilot sample was drawn from different regions and urban/rural locations. The aim was not for the pilot sample to be representative of the full sample, but rather: (i) for selected schools and students to reach the extremes of expected performance in each subject to identify potential 'floor' and ceiling' effects; and (ii) to cover all languages of administration, to investigate translation suitability.

The Wave 1 pilot sample consisted of 1,248 students attending a total of 13 schools in six regions: Oromia, Tigray, Somali, SNNP, Afar and Addis Ababa. The Wave 2 pilot sample consisted of 835 students attending a total of nine schools in five regions: Oromia, Tigray, Somali, SNNP and Addis Ababa. The Wave 2 pilot sample also included 36 teachers at the same schools, who completed teacher psychosocial scales, and 35 teacher trainees at Kotebe University College who completed the teacher professional knowledge questionnaire (this questionnaire was also completed by 18 of the 36 teachers at schools – those who teach mathematics).

Based on pilot data analysis, qualitative feedback and translation checks, items on tests, in background questionnaires, in psychosocial scales and on teacher professional knowledge questionnaires were reviewed for removal, revision and/or inclusion in final instruments.[33]

4.3 Survey administration

Both waves of the school survey were conducted by trained fieldworkers from the Ethiopian Development Research Institute (EDRI). Before each wave of data collection, fieldworkers received full training on the instruments to be used in the survey. Fieldworker training was delivered jointly by Young Lives team members from the University of Oxford and EDRI.

In both Wave 1 and Wave 2 of the survey, each fieldwork team spent between two and five days in each school, depending on the number of students enrolled. In order to avoid fatigue, the mathematics and English tests could not be administered on the same day. Wave 1 fieldwork began in October 2016 and was completed within one month. Wave 2 fieldwork began in May 2017 and was also completed within one month.

Test data were entered by EDRI. Quality control checks of test and questionnaire data were completed manually once data had been entered.

Survey administration, including securing permission to visit schools, making training arrangements, and dealing with fieldwork logistics, was overseen by EDRI.

33 See Azubuike et al. (2017) and Moore and Rossiter (forthcoming) for further information on piloting procedures for the different instruments.

References

Alemu, D., and A. Tekleselassie. (2011) 'Comparative Analysis of Instructional Language Issues in Ethiopia and the United States', *Creative Education* 2.4: 402–7. https://file.scirp.org/pdf/CE20110400012_46965270.pdf (accessed 20 September 2017).

Azubuike, O.B., R. Moore, and P. Iyer (2017) *Young Lives School Surveys, 2016–17: The Design and Development of Cross-Country Maths and English Tests in Ethiopia, India and Vietnam*, Young Lives Technical Note 39, Oxford: Young Lives.

Graddol, D. (2010) *English Next: India*. London: British Council.

Grønmo, L.S., M. Lindquist, A. Arora and I.V.S Mullis (2015) *TIMSS 2015 Mathematics Framework.* Boston, MA: TIMSS and PIRLS International Study Centre.

Kane, T., and D. Staiger (2008) *Estimating teacher impacts on student achievement: An experimental evaluation*, NBER Working Paper 14607, Cambridge, MA: National Bureau of Economic Research.

Little, A., and O.B. Azubuike (2017) *Young Lives School Surveys, 2016–17. The Development of Non-Cognitive Instruments in Ethiopia, India and Vietnam*, Young Lives Technical Note, Oxford: Young Lives.

Ministry of Education (2009) *Education Statistics Annual Abstract 2001 E.C. (2008/09)*, Addis Ababa: Government of Ethiopia.

Ministry of Education (2013) *Ethiopian 4th National Learning Assessment of Grades 4 and 8 Students, Data Analytic Report*, Addis Ababa: Government of Ethiopia.

Moore, R. and Rossiter, J. (forthcoming) *Young Lives School Surveys, 2016-17. The Design and Development of Teacher Measures for use in Ethiopia, India and Vietnam*, Young Lives Technical Note, Oxford: Young Lives.

OFQUAL (2011) *Functional Skills Criteria for English. Entry 1, Entry 2, Entry 3, Level 1 and Level 2*, Coventry: OFQUAL.

Outes-Leon, I. and A. Sanchez (2008) *An Assessment of the Young Lives Sampling Approach in Ethiopia*, Young Lives Technical Note 1. Oxford: Young Lives.

Rankin, K., D. Cameron, K. Ingraham, A. Mishra, J. Burke, M. Picon, J. Miranda, and A. Brown (2015) *Youth and transferable skills: an evidence gap map, 3ie Evidence Gap Report 2*. New Delhi: International Initiative for Impact Evaluation (3ie).

Tafere, Y. (2014) *Educational Aspirations and Barriers to Achievement for Young People in Ethiopia*, Young Lives Working Paper 120, Oxford: Young Lives.

Van der Linden, W., and R. Hambleton (eds.) (2013) *Handbook of modern item response theory*, New York: Springer Science and Business Media.

Vjucich, D. (2013) *Policy and Practice on Language of Instruction in Ethiopian Schools: Findings from the Young Lives School Survey*, Young Lives Working Paper 108, Oxford: Young Lives.

Appendix 1. Site descriptions

Region	Cluster ID	Reference in text	Description of site
Addis Ababa	1	AD1	An overcrowded area in the centre of the capital city.
Addis Ababa	2	AD2	An industrial area in the southern part of the capital city.
Addis Ababa	3	AD3	A slum area in the capital city.
Amhara	4	AM4	A tourist town with some extremely poor neighbourhoods.
Amhara	5	AM5	A poor rural community
Amhara	6	AM6	A rural area near Lake Tana.
Amhara	7	AM7	A rural food-insecure area.
Oromia	8	OR8	A rural area near lake Ziway.
Oromia	9	OR9	A drought-prone rural area.
Oromia	10	OR10	A fast-growing town.
Oromia	11	OR11	A relatively rich rural area in the outskirts of Debrezeit town.
SNNP	12	SN12	A densely populated rural area growing enset ('false banana').
SNNP	13	SN13	A densely populated town
SNNP	14	SN14	A fast-growing business and tourist town.
SNNP	15	SN15	A coffee-growing rural area.
SNNP	16	SN16	A poor and densely populated rural community.
Tigray	17	TI17	A drought-prone rural area highly dependent on government support.
Tigray	18	TI18	An extremely poor rural area dependent on the Productive Safety Net Program and other government support.
Tigray	19	TI19	A small, very poor town.
Tigray	20	TI20	A model rural area known for its success in soil and water conservation.
Somali	21	SO21	A drought-prone area where animal husbandry is the main means of livelihood for the community.
Somali	22	SO22	A drought-prone area where animal husbandry is the main means of livelihood for the community.
Somali	23	SO23	An area within the regional capital, Jijiga. Compared with other sites, the economy is stronger, consisting of trade, services, business and government employment.
Somali	24	SO24	A drought-prone rural area affected by frequent shortages of water and grazing land. The main means of livelihood for the local community are animal husbandry and farming.
Afar	25	AF25	A town about 700km from the capital city, Addis Ababa. A small power station, a health centre and various primary and secondary schools have been constructed since 2005.
Afar	26	AF26	A better-off rural area where most households own livestock.
Afar	27	AF27	A drought-prone rural area affected by frequent shortages of water.
Afar	28	AF28	A drought-prone urban area affected by frequent shortages of water.
Afar	29	AF29	A drought-prone rural area affected by frequent shortages of water.
Afar	30	AF30	A small urban town densely populated by commercial farm workers and government employees.

Young Lives School Survey, 2016-17: Evidence from Ethiopia

This report gives an overview of the Young Lives school effectiveness survey conducted at the beginning and end of the 2016-17 (2009 E.C.) academic year. It provides a descriptive summary of the data collected from 12,182 students in Grade 7 and Grade 8, learning in 30 sites across seven of Ethiopia's eleven regional states and city administrations.

The school effectiveness survey was designed to allow analysis of what shapes children's learning and progression over a school year. The data will allow researchers to understand, describe and explain school and education system effectiveness. The survey focuses on issues of attainment (e.g. progression and grade completion) and achievement (e.g. on the learning and other outcomes delivered by the school system) in a structure that links students to teachers, classrooms and schools.

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 4 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 4 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