RISE
RESEARCH ON IMPROVING SYSTEMS OF EDUCATION

INSIGHTS

Trends in Mathematics Learning in Ethiopia: 2012 - 2019

by Dawit T. Tiruneh, Ricardo Sabates, Caine Rolleston, and John Hoddinott

RISE Ethiopia Country Research Team

Introduction

Ethiopia is among the lowest-income countries but has greatly increased funding for the education sector over the past two decades. In 2016-17, education accounted for 27 percent of total government expenditure, which is significantly higher than the government's commitment to internationally agreed targets of 20 percent of the national budget for education (UNICEF, 2017). International development agencies have also been calling for greater resources to be devoted to education and have increased their levels of assistance for education projects in Ethiopia (Ministry of Education, 2015; World Bank, 2017).

In line with Sustainable Development Goals (SDG) 4.1 and 4.2, education investment in Ethiopia has focused on increasing pre-primary and primary school enrolment, while another has been on improving learning outcomes equitably for all. Accordingly, pre-primary gross enrolment has rapidly expanded from less than 300,000 pupils in 2008-09 (4 percent), to over 3.5 million (45 percent) in 2019-20; and primary education enrolment from 3 million learners in the early 1990s to over 20 million in 2019-20 (Ministry of Education, 2020). However, despite the tremendous progress in expanding access to pre-primary and primary education, learning levels have remained low or declined (Ministry of Education, 2010, 2015; World Bank, 2017). A large share of children complete their primary education lacking basic literacy and numeracy skills (e.g., NEAEA, 2016; Tiruneh, Hoddinott, et al., 2021; USAID, 2019).

In this Insight Note, we explore the possible explanations for the decline in learning levels among primary school pupils in relation to the General Education Quality Improvement Programme (GEQIP) reforms that were

Key Points

- This Insight Note reports trends in mathematics learning for Grade 4 pupils in Ethiopia, based on a longitudinal survey of 33 schools between 2012-13 and 2018-19.
- Despite the implementation of the General Education Quality Improvement Programme (GEQIP) Phase II reforms, pupils' mathematics learning levels declined between 2012-13 and 2018-19.
- Progress in mathematics in the 2018-19 academic year improved slightly compared to 2012-13, but there is a difference in magnitude of learning progress for the two periods between pupils across rural-urban locations, regional states, and family economic backgrounds.
- There is an overall improvement in measures of school infrastructure and in teacher qualifications between 2012 and 2018, and there is some evidence of changes in student composition between the two periods.
- Consistent with the GEQIP-II reforms in terms of supporting access and retention, pupils in 2018-19 were more likely to: have attended pre-school; have lower levels of absenteeism; and have a lower dropout rate, compared with pupils in the same grade in 2012-13.
- Compared with the students in 2012, those in the 2018 cohort had caregivers that were less likely to be literate and had fewer assets at home. Pupils were also travelling a relatively longer distance to school, and were slightly older compared with those in the 2012 group.
- Differences in mathematics learning levels and learning progress between disadvantaged pupils (i.e., those from rural areas, emerging regions, and the lowest socio-economic backgrounds) and their relatively advantaged counterparts are discussed in relation to the GEQIP-II reforms.

intended to improve quality and equity in the Ethiopian basic education system. We examine the extent to which mathematics learning levels for Grade 4 pupils have declined over time, despite the implementation of reforms to improve them, as well as the lessons that may be drawn from this. We also examine whether there is any difference in the benefits of the educational reforms for pupils from disadvantaged backgrounds (i.e., from rural areas, emerging regions, and from the lowest socio-economic background). We make use of a unique longitudinal dataset on 33 schools in six regions of Ethiopia covering the period 2012 to 2019.

Educational Reforms in Ethiopia: The General Education Quality Improvement Programme (GEQIP)

Recognising the inadequacy of the primary education system to equip children with the required knowledge and skills, Ethiopia began to undertake major efforts in 2008 to raise learning outcomes equitably through the introduction of government- and donor-supported programmes. In the latter category, one of the most prominent programmes is the set of GEQIP reforms (World Bank, 2008, 2015). These reforms have been implemented in Ethiopia since 2008 in three consecutive phases: GEQIP-I (2008-2012); GEQIP-II (2012-2018); and currently GEQIP for Equity (GEQIP-E: 2018-2022). The reforms have been comprehensive and nationwide, and their overall aim is to enhance pupils' learning outcomes equitably by improving teaching and learning conditions in schools, and to strengthen educational institutions and service delivery at the federal and regional level (World Bank, 2008, 2017). GEQIP-I and GEQIP-II reforms focused on providing essential inputs to all public schools to improve teaching and learning, including: building additional classrooms; increasing the supply of qualified primary school teachers; providing continuous in-service training for teachers to enhance their content knowledge and pedagogical content knowledge; providing pupils with textbooks for each subject; and funding school improvement plans through per capita school grants based on enrolment (World Bank, 2008; 2012). The major difference in GEQIP-II was the inclusion of information and communication technology as an additional component.

Although the implementation of GEQIP-I and GEQIP-II reforms was completed some years ago, large-scale longitudinal studies to evaluate their impact have not been conducted to date. In particular, the reforms' effects on the most marginalised pupils, including girls, those from the lowest income families, from rural locations and so-called "emerging" regions are not well understood. The Ministry of Education (MoE) in Ethiopia conducted national-level assessment studies focusing on mathematics and reading comprehension with Grade 4 and Grade 8 children in 2011 and 2015, parallel to the implementation of the GEQIP-I and GEQIP-II reforms, respectively. The findings indicated that average mathematics and reading comprehension scores for these grades were below the minimum expected standards set by the MoE (Ministry of Education, 2015; NEAEA, 2016; USAID, 2019). Although these and other similar findings from the national learning assessments are useful, the studies use cross-sectional data, which are less suited to assessing the contributions of GEQIP-I and GEQIP-II reforms to children's learning progress over time.

The objective of the study

In this paper, we examine the extent to which school-based GEQIP-II reforms were followed by improvements in pupils' learning progress over time. We particularly focus on comparing the learning levels and learning progress over time between pupils in rural areas and their urban counterparts, and between those from the lowest family socio-economic background and their relatively wealthier peers. The study employs unique longitudinal large-scale data collected in Ethiopia at the start and end of the GEQIP-II reforms, targeting the same schools in six regions. The data includes repeated measures of learning outcomes, as well as relevant associated information such as the child's background, teacher training, and school characteristics.

At the start of GEQIP-II, it was established, among other things, that (a) textbooks were not available on time and in sufficient quantities in primary schools; (b) the provision of relevant school facilities such as a functional library, pedagogical resource centres, computers and internet access was insufficient; and (c) poor teacher training was a major contributor to low pupil learning outcomes (World Bank, 2013). It was assumed that if these and other related inputs/resources were provided to primary schools, the pupils' learning levels would improve. Although several cross-sectional studies have been conducted in Ethiopia over the past decade to examine pupils' levels of performance, no formal evaluations have been made to assess the contributions of the GEQIP-II reforms to equitable learning progress over time, in combination with other relevant associated characteristics. In particular, the evidence is scant on the

benefits of educational reforms in Ethiopia to the most disadvantaged. We therefore address the following research questions:

1. How have mathematics learning outcomes among Grade 4 pupils in Ethiopia changed over the period 2012 and 2018?
2. Are there differences in mathematics learning levels and learning progress between disadvantaged pupils (i.e., from rural areas, emerging regions, and lowest socio-economic backgrounds) and their relatively advantaged counterparts for the same period?

Data and Methods

Data

We used data from the Young Lives Ethiopia (YL) 2012-13 School Survey and the RISE Ethiopia 2018-19 Household and School Surveys. Young Lives is a longitudinal study of childhood poverty conducted in Ethiopia starting from 2002, tracing the lives of children through school surveys. The YL 2012-13 School Survey included nearly 12,000 pupils studying in both Grades 4 and 5, in purposely selected sites located across seven regions in Ethiopia: Addis Ababa, Afar, Amhara, Oromia, Southern Nations, Nationalities and People's (SNNP), Somali, and Tigray. The school survey offers a unique perspective on regional and site differences across child, teacher and school characteristics, and the factors influencing progress in mathematics and reading over the course of a single school year.

The YL 2012-13 survey was conducted in two waves: Wave 1 at the start of the 2012-13 school year in October 2012, and Wave 2 towards the end of the school year in May 2013. In Wave 1, the survey included a pupil questionnaire; an assessment of mathematics and reading comprehension; an assessment of teacher content knowledge; and a principal questionnaire to gather some indicators of school and class quality. In Wave 2, pupils completed a second set of learning assessments in mathematics and reading comprehension. A total of 10,068 pupils in 94 schools and 280 Grade 4 and Grade 5 classes were surveyed in both Wave 1 and Wave 2 (for details, see Aurino et al., 2014). For this study, we have used data which focuses only on Grade 4 pupils and their associated characteristics (household, teacher and school levels) to analyse the determinants of progress in maths over a school year.

RISE Ethiopia has adopted a longitudinal design that is very similar to the YL survey (see, Tiruneh, Sabates, et al., 2021). The target population of the RISE Ethiopia 2018-19 household and school surveys included Grade 1 and Grade 4 school children from 166 schools, their parents (or primary caregivers), school principals, and Grade 4 mathematics and reading teachers distributed across seven regions: Addis Ababa, Amhara, Benishangul Gumuz (Be-Gu), Oromia, SNNP, Somali and Tigray. The survey and sample design have been described in Hoddinott, Iyer, Sabates, and Woldehanna (2019). Similar to YL, the 2018-19 RISE Ethiopia surveys were conducted in two waves: Wave 1 at the start of the 2018-19 school year in November, and Wave 2 towards the end of the 2018-19 school year in June. Wave 1, included both the school and household surveys. In Wave 2, pupils completed a second set of learning assessments in mathematics and reading comprehension, and teachers completed a questionnaire and an assessment of their mathematics content knowledge. The RISE Ethiopia data employed in this study focuses on Grade 4 children only.

Sampling

Among the total 166 schools in the RISE Ethiopia 2018-19 surveys (hereafter, 2018-19), the sampling strategy included 33 schools that had also participated in the YL 2012-13 school survey (hereafter, 2012-13). These overlapping schools were selected from six regions (excluding Benishangul-Gumuz region). The 33 schools in the 2012-13 and 2018-19 surveys enable us to explore how GEQIP-II's reform indicators of school resources (e.g., pupil-textbook ratio, access to key educational inputs, etc.) and pedagogical supplies (e.g., teacher content knowledge, teachers' qualification, teachers' experience, etc.) changed between 2012 and 2019.

For the 2012-13, all Grade 4 pupils from the 33 schools were selected as participants. However, for the 2018-19 survey, only two Grade 4 classes were randomly selected from each of the 33 sample schools. At a pre-survey tracking exercise, 28 pupils were randomly identified from up to two Grade 4 classes in each sample school. Because the number of Grade 4 classes varies across schools, and as the sample of 28 pupils per school for the 2018-19 survey was drawn

from classes that vary in student population, we weighted the observations taking into account the number of Grade 4 classes and the number of Grade 4 students per school. In effect, the weights reflect both the selection of Grade 4 classes and the probability of a Grade 4 child being included in the sample, determined largely by school size (pupils in smaller schools have a higher probability of being included). We have estimated the key descriptive statistics using these weights to adjust for sampling effects. We compared the average scores from the YL 2012-13 with the RISE 2018-19 weighted scores.

Attrition analysis

Table 1 shows that 2,652 Grade 4 pupils participated in the baseline surveys for YL 2012-13, and 816 pupils for the RISE 2018-19. The attrition rate between baseline and endline for the YL 2012-13 sample was nearly 17 percent, while it was nearly 16 percent for the RISE 2018-19 sample. Although the total RISE sample expected at baseline from the 33 schools was 924 (33 schools x 28 pupils per school), it should be noted that the actual number of pupils available per school at the baseline was less than 28 pupils in some of the sample schools due to student absenteeism. For the YL 2012-13, we have 2,190 pupils who had both baseline and endline scores, and for the RISE 2018-19, we have 689 pupils who took both the baseline and endline tests. The main reasons for attrition between the baseline and endline for both surveys were: class absenteeism at the time of test administration; dropouts; a change of school because parents relocated to other areas; and a failure to track some pupils due to lack of proper class rosters in the surveyed schools.

To avoid non-response bias arising due to high attrition rates for both YL 2012-13 and RISE 2018-19, we predicted endline test scores for attriters who were absent at the endline, using a regression model.[1] On average, these pupils were lower scorers than those retained in the sample for both YL 2012-13 and RISE 2018-19. We used the model predictions to impute an endline test score for attriters and we were then able to compare the mean score at endline with and without these imputed values. Including the imputations reduced the mean endline score slightly, as expected, given that attriters are both lower performing and more disadvantaged pupils. The change in mean endline scores is small enough that it does not alter the interpretations of the findings of this paper.

Table 1. Young Lives 2012-13 and RISE 2018-19 sample size and attrition rate by region

Region	Number of schools	YL 2012-13 sample size			RISE 2018-19 sample size		
		Baseline	Traced in Endline	Attrition rate (%)	Baseline	Traced in Endline	Attrition rate (%)
Addis Ababa	3	260	220	15.4	83	79	4.8
Amhara	6	513	427	16.8	159	141	11.3
Oromia	6	505	409	19.0	157	122	22.3
SNNP	7	674	550	18.4	177	143	19.2
Somali	4	191	137	28.3	62	49	21.0
Tigray	7	509	447	12.2	178	155	12.9
Total	**33**	**2,652**	**2,190**	**17.4**	**816**	**689**	**15.6**

Source: Young Lives 2012-13 and RISE Ethiopia 2018-19

It should be noted that the sites in the six regions in the 2012-13 survey were selected purposively with a pro-poor bias, i.e., with a focus on sites that are food-insecure and drought-prone (for details, see Aurino et al., 2014). Thus, the

[1] Independent variables included in the regression model to predict endline scores are baseline average score, gender, family economic background, preschool attendance, and primary caregivers' literacy.

findings presented in this paper based on the 33 common schools between the two surveys are not representative of the impacts of GEQIP-II at regional or national levels because the school selection was not random or representative while GEQIP-II was implemented across the country. However, the 33 common schools provide a unique panel dataset with comparable test scores and a rich set of indicators at the pupil, teacher and school level. This allows for robust analysis of trends in learning outcomes across location, regional states, and economic background, and their potential causes and consequences, including in relation to the GEQIP-II reforms as implemented in these 33 schools at least.

Instruments

Mathematics test for Grade 4 pupils

Comparable mathematics tests were administered both at the start and end of the school year for both the 2012-13 and 2018-19 cohorts in order to measure pupils' learning outcomes in the subject. The 2012-13 survey included 25 multiple-choice items in each wave (i.e., at the start [baseline] and end [endline] of the 2012-13 school year). The endline test included 19 common (anchor) items from the baseline test and six that were unique. The 2018-19 tests were adapted from the 2012-13 versions. The baseline test contained 25 items, and were administered at the start of the 2018-19 academic year. The endline test, administered at the end of the 2018-19 academic year, included 15 common (anchor) items from the baseline test and 10 items that were unique. Taking both the 2012-13 and 2018-19 school surveys together, there were 13 items common across the four waves. In total, there were 41unique items, and our analysis of item fit indicated that the items overall functioned well across the four waves with acceptable item difficulty and item discrimination indices.

Teacher mathematics content knowledge test

In the 2012-13 survey, 30 content knowledge test items were administered to Grade 4 mathematics teachers. Twenty of the same items were administered to Grade 4 mathematics teachers in the 2018-19 survey. All the items in both surveys functioned well, with acceptable item difficulty and item discrimination indices.

School principal and teacher questionnaires

School principal and teacher questionnaires were administered in the common sample schools in both the 2012-13 and 2018-19 periods. The principal questionnaires in both rounds focused on gathering information on the school principals' levels of education, and their experience, as well as on indicators of school quality, for example: pupils' access to educational resources (library, textbooks, computers, radio, working toilets, access to school grants, etc.), the and provision of continuous professional development (CPD) opportunities for teachers. The teacher questionnaire focused on gathering data related to the teacher's age, experience, education levels, and teacher training qualifications.

Item Response Theory Modelling

To enable comparable estimations of learning levels and progress in mathematics over the school year between the 2012-13 and 2018-19 cohorts, and to compare teachers' mathematics content knowledge over time, we employed a concurrent calibration approach in an Item Response Theory (IRT) modelling framework. A two parameter-logistic IRT model (2PL IRT) was fitted to the item responses. The 2PL IRT model provides parameter estimates on a common interval scale. In concurrent calibration, item parameters are estimated simultaneously using pooled data from all waves, with responses to the items that were unique to each group treated as missing for respondents that did not receive them. The anchor items provide the link between tests, while the unique items increase the precision of estimates for individual tests. This approach has proven to be effective in accurately estimating item parameters for all the test takers, especially when we link scores across time. Following our earlier work (Rolleston et al., 2013; Tiruneh, Hoddinott, et al., 2021), we transform the pupils' latent trait estimates for the entire pooled sample to a scale with a mean of 500 and a standard deviation (SD) of 100, for ease of reference. Similarly, IRT methods were employed to calibrate a common scale metric for teacher knowledge items across both YL and RISE surveys.

Findings

Learning level and learning gain for the common sample in 2012-13 and 2018-19

Table 2 presents the Grade 4 mathematics mean scores[2] and their standard deviations for the 2012-13 and 2018-19 common school sample, by region, locality and socio-economic background.

Table 2. Learning level and learning gain for the 2012-13 and 2018-19 cohorts, by region, locality and socio-economic background

			2012-13				2018-19			
Total		N[1]	Baseline: Mean (SD)	Endline: Mean (SD)	Gain	N	Baseline: Mean (SD)	Endline: Mean (SD)	Gain	Decline in baseline scores between 2012 & 2018 in SD
		2,190	493 (84)	524 (97)	31***	689	451 (89)	488 (97)	37***	.42
Region	Addis Ababa	220	524	568	44***	79	508	544	36***	.16
	Amhara	427	502	536	34***	141	472	501	29***	.30
	Oromia	409	484	530	46***	122	425	489	64***	.59
	SNNP	550	503	529	26***	143	439	468	29***	.64
	Somali	137	461	488	27***	49	417	446	29***	.44
	Tigray	447	474	490	16***	155	463	501	38***	.11
Locality	Urban	1,325	503	537	34***	308	464	499	35***	.39
	Rural	865	477	503	26***	381	435	474	39***	.42
	Difference		27***	34***			29***	25***		
Family economic	*Wealthiest*	809	508	542	34***	148	463	493	30***	.45
	Poorest	772	477	504	27***	346	441	486	45***	.36
	Difference		31***	38***			22***	7		

Note: t-test of the maths mean gain is significant at ***p<0.001; **p<0.05.

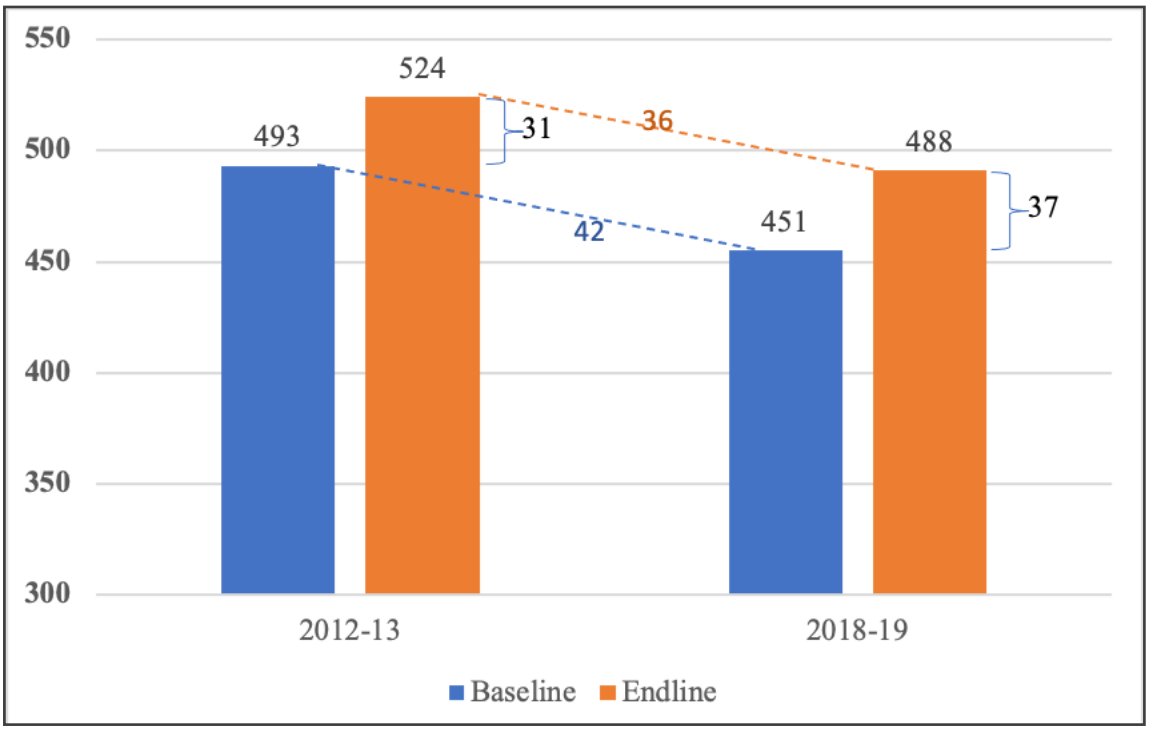
Table 2 contains a considerable amount of information. We begin with broad trends, then turn to a more detailed examination of these data.

We start with the baseline means (results from the mathematics test administered at the start of the school year) for the 2012-13 and 2018-19 samples. The striking result is that this mean falls from 493 in 2012 to 451 in 2018, a 0.42 SD drop. In both samples, learning occurs over the course of Grade 4 as evidenced by increases in end-of-year test scores: gains of 31 and 37 points (0.31SD and 0.37 SD) in 2012-13 and 2018-19 respectively. Put differently, with yearly learning gains

[2] The mean score is an interval scale centred on 500, with 500 defined as the mean of the pooled sample - from 2012-13 and 2018-19 cohorts.

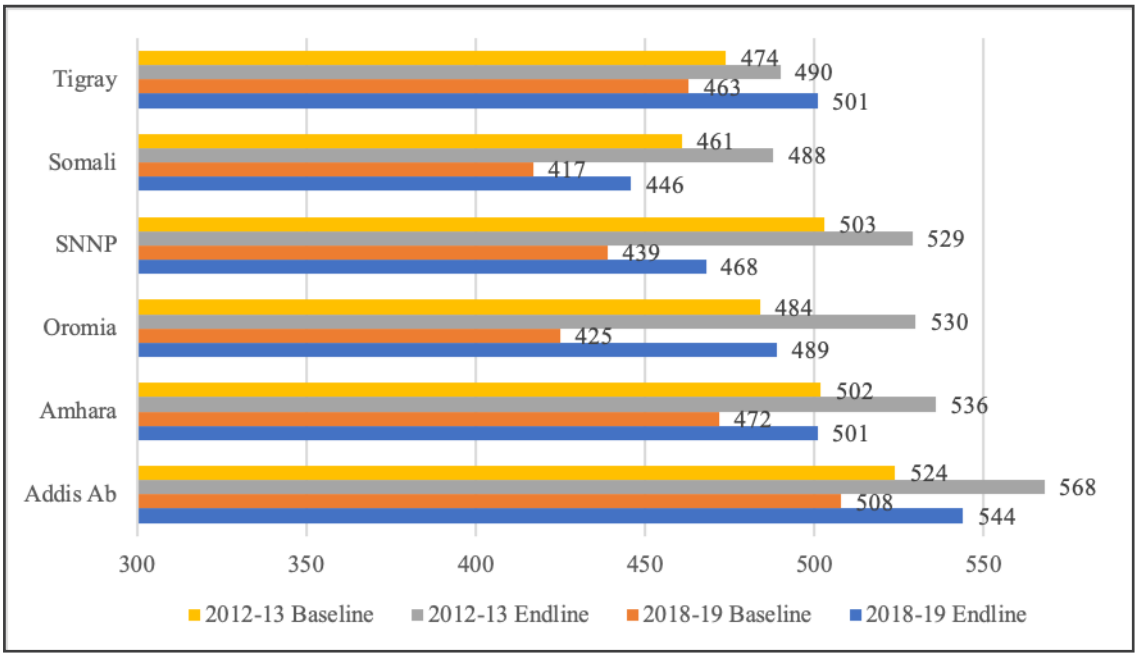
falling somewhere between 0.31-0.38 SD, the 2012-2018 drop of 0.42 SD in the start-of-school scores is equivalent to one year of instruction in mathematics (Figure 1). Furthermore, the conjunction of much lower start-of-school scores in 2018-19 and the slightly larger gains over the 2018-19 school year (compared to those in 2012-13) means that the end-of-year scores in 2018-19 were lower, not only when compared to end of the school year scores in 2012-13, but also when compared to the beginning of the 2012-13 school year scores.

Figure 1. Decline in mathematics achievement for the 2012-13 and 2018-19 common sample



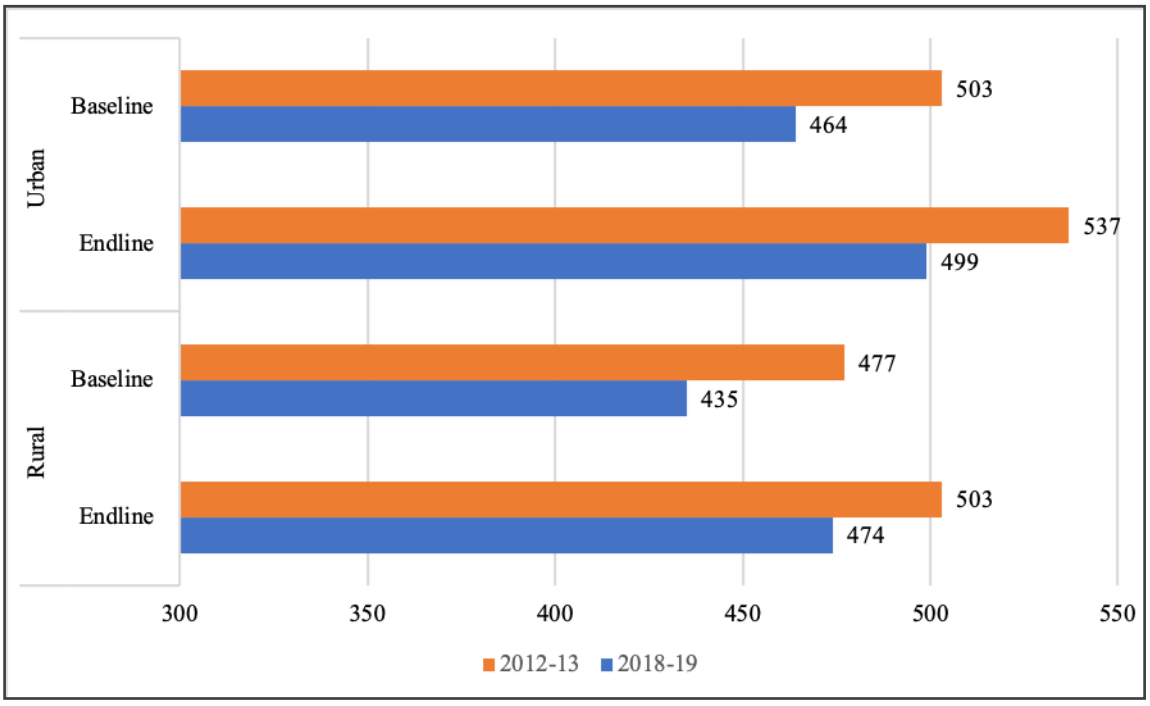
Next, we consider three disaggregations: by region, by rural/urban location, and by household wealth. We begin with region. Start-of-school test scores decline in all regions between 2012 and 2018, but the magnitude of these declines differ markedly. They fall by 0.30SD, 0.16SD, and 0.11SD in Amhara, Addis Ababa, and Tigray regions respectively, and by the even greater magnitudes of 0.64SD, 0.59SD, and 0.44SD in SNNP, Oromia and Somali respectively. Learning gains were slightly higher in 2018-19 than in 2012-13 in all regions except Addis Ababa and Amhara. However, once again, there is considerable variation in learning gains across those regions, as shown in Figure 2: learning gains increased in Oromia (from 0.46SD in 2012-13 to 0.64SD in 2018-19), SNNP (from 0.26SD to 0.29SD), Somali (from 0.27SD to 0.29SD), and Tigray (from 0.16SD to 0.38SD), but fell by 0.08SD in Addis Ababa and 0.05SD in Amhara region. Overall, during the period in which the implementation of GEQIP-II reforms occurred, regional inequality in learning levels widened substantially over the six-year period, either because of differences in the size of the decline in start-of-school-year test scores, or because of changes in the magnitudes of learning gains, or both.

Figure 2. Decline in mathematics learning levels over time, by region



In terms of rural-urban location, again, pupils from both locations made significant progress over the school year (in both 2012-13 and 2018-19). As shown in Figure 3, the progress in mathematics over a school year for urban pupils in 2018-19 was similar to that of their counterparts in 2012-13 (nearly a third of standard deviation). The progress for rural pupils in 2018-19, however, was significantly higher (39 points) compared to the progress made by their counterparts in 2012-13 (26 points).

Figure 3. Decline in mathematics learning levels over time, by rural-urban location



It should be noted again that there is an overall decline in average mathematics learning levels between 2012 and 2018 for both the urban and rural cohorts. The most notable is that the average end of school year mathematics score for rural pupils in 2018-19 (474 points) was lower than the average start of school year score for Grade 4 rural pupils in 2012-13 (477 points). This means that despite the strong progress in mathematics in the 2018-19 school year (39 points), after one year in school, Grade 4 rural pupils in 2018-19 did not reach the level that rural Grade 4 students achieved in 2012-13 before starting the school year. However, the urban-rural learning level gap decreased significantly between 2013 and 2019. The average gap at the end of Grade 4 in 2013 was 34 points, but it decreased to 25 points in 2019.

Lastly, we constructed a wealth index.[3] Having done so, we divided our sample into tertiles and reported results for children in households in the wealthiest and poorest tertiles. In both 2012 and 2018, start-of-year scores are lower for children residing in households in the poorest wealth tertile. This wealth gap in mathematics average scores shrinks slightly between 2012 and 2018. In 2018-19 the poorest group made significantly higher progress (45 points) over a school year, compared to their wealthiest counterparts (30 points). Consequently, the gap in mathematics learning levels at the end of Grade 4 between the relatively wealthiest and poorest group declined markedly from 38 points in 2013 to 7 points in 2019.

There may be several possible explanations for these findings in the common school sample over the six-year period in question. In the following sections, we discuss all the findings in relation to pupils' backgrounds and GEQIP-II related school-resources and teacher-quality indicators.

Key pupils' background factors in 2012-13 and 2018-19: Rural and urban cohorts

Table 3 below provides descriptive statistics for key background indicators between the 2012-13 and 2018-19 cohorts, first for the overall sample, followed by rural-urban localities. Overall, pupils in the 2018-19 sample appear more disadvantaged in that: (i) their caregivers are less likely to be literate; (ii) they have fewer household assets; (iii) they travel relatively longer distances to school; and (iv) they are slightly older. These trends suggest that over time, school enrolment has reached a greater proportion of pupils overall, thereby including more relatively disadvantaged pupils in terms of their household assets and backgrounds, i.e., groups of pupils who might not have enrolled in the 2012-13 school year. When we look at the rural vs. urban sample, the results with respect to some of the indicators are mixed. Although the overall trend shows a decline, the proportion of literate primary caregivers for urban pupils has increased slightly (2 percentage points), and these pupils actually travelled shorter distances to get to school. The differences are statistically significant. Thus, it is actually rural pupils who exhibit increased disadvantages over time, as measured by caregivers' literacy and the commuting distance to school.

Despite being more disadvantaged on average in wealth terms, pupils in 2018-19 benefited from supply-side policies and improvements: they were (i) more likely to have attended pre-school; (ii) absent from school less frequently; and (iii) less likely to have dropped out (perhaps related to "automatic promotion"). The trend is the same when we look closely at pupils in rural and urban localities. While we cannot make strong statements about causality, it is important to note that these improvements are consistent with the GEQIP-II reforms, in terms of supporting access, and pupil retention.

When we look specifically into household assets, patterns vary somewhat by region, as illustrated in Figure 4. Tigray region—where declines in start-of-school scores were modest—saw the lowest decline in pupils' background indicators in the form of household assets, while Somali region (where declines in start-of-school scores were much higher) saw the steepest decline between 2012 and 2018.

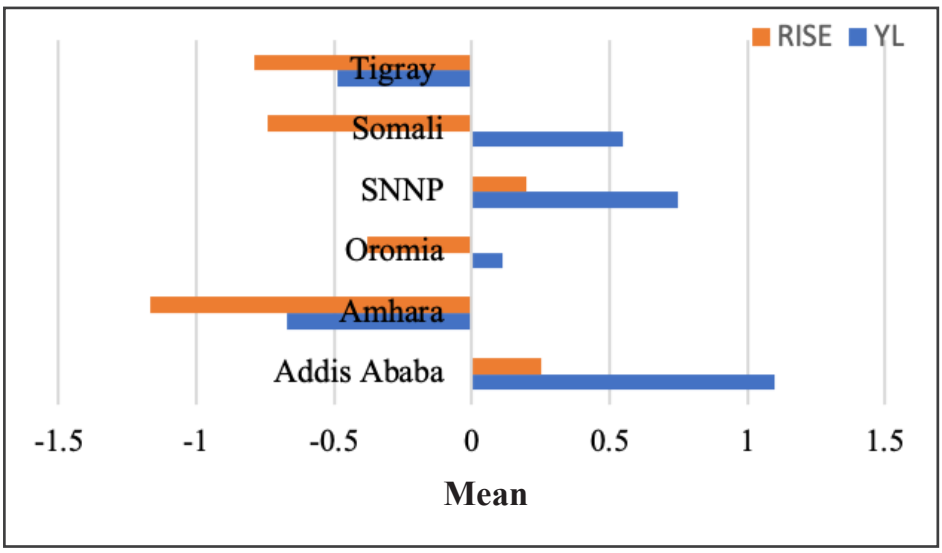
[3] The wealth index was constructed based on households' durable assets, which serves as a proxy measure for overall household economic advantage. We decided to exclude indicators related to access to electricity, access to tap water and other services because they are not applicable to the rural parts of Ethiopia.

Table 3. Key pupil background indicators for the 2012-13 and 2018-19 common school cohorts, by rural-urban location

Indicator	2012-13				2018-19			
	Overall (n=2,190)	Rural (n=865)	Urban		Overall (n=689)	Rural (n=381)	Urban (n=308)	
Average student age	11.04	11.05	11.03	-0.02	11.13	11.4	10.8	-0.6***
Proportion of students who attended preschool, %	42.0	21.0	55.0	34.0***	50.0	36.0	66.0	30.0***
Proportion of students ever dropped out before Grade 4, %	19.0	20.0	18.0	-2.0	11.0	10.0	11.0	1.0
Average number of absent days in the current school year	1.64	2.44	1.11	-1.33***	1.46	2.03	0.71	-1.32***
Household durable assets[4], average	0.12	-0.72	0.67	1.39***	-0.47	-1.00	0.20	1.20***
Primary caregivers' literacy, %	50.0	45.0	52.0	7.0**	41.0	31.0	54.0	23.0***
Average time taken to walk to school (in minutes)	18.35	20.22	17.15	-3.07***	21.84	26.59	16.0	-10.56***

Note: t-test of the differences is significant at ***p<0.001; **p<0.05

Figure 4. Household assets by region for 2012-13 and 2018-19 common sample



[4] Households' durable assets were used to measure household economic status, which serves as a proxy measure for overall household economic advantage. We decided to exclude items related to access to electricity, access to tap water and other services because they do not apply to rural areas of Ethiopia.

Indicators for School-Resources and Teacher-Quality

Across the 2012-13 and 2018-19 cohorts, Table 4 reports the indicators of school resources for rural and urban schools, focusing on those related to GEQIP-II reforms. Table 5 presents selected background characteristics of teachers included in both the 2012-13 and 2018-19 surveys, for rural and urban schools. At the school level, the pattern of change for the overall sample is somewhat mixed, but broadly positive, with some improvement in the availability of: (i) a pedagogical resource centre; (ii) computers and internet access; and (iii) full-day shift schooling being recorded, while fewer schools reported receiving money from a school grant. Looking at the rural-urban differences in key school resources, the pattern is similar. The proportion of functional libraries, working computers, functional pedagogical resource centres, and separate toilets for girls and boys increased for both rural and urban schools in each of the two cohorts. Rural schools showed better improvement across some indicators than their urban counterparts, including the number of additional classrooms built, access to working radios for instructional purposes, and the proportion of schools operating a full-day shift.

Table 4. School resource indicators related to GEQIP-II reforms, by rural-urban location

		2012-13			2018-19		
	Indicator	Rural	Urban	Diff.	Rural	Urban	Diff.
1.	Schools that provide one G4 maths textbook per student, %	63.0	85.0	22.0	58.0	64.0	6.0
2.	School received funding from school grant last year, %	95.0	93.0	-2.0	84.0	71.0	-13.0
3.	Average number of classrooms in the school	8.95	20.0	11.05***	11.05	19.71	8.66**
4.	Average number of students per class	51.0	56.0	5.0	50.0	53.0	3.0
5.	Presence of a functional library, %	58.0	86.0	28.0*	68.0	100.0	32.0**
6.	Average number of working computers used by the students	0.42	2.0	1.58	0.68	5.07	4.39**
7.	Schools with access to functional Internet for the students, %	0	7.0	7.0	0	29.0	27.0**
8.	Schools with access to working radios, %	53.0	86.0	33.0**	74.0	79.0	5.0
9.	Schools with a functional pedagogical resource centre, %	47.0	64.0	17.0	79.0	93.0	14.0
10.	Schools operating a full-day shift, %	0	21.0	21.0**	11.0	21.0	10.0
11.	Average number of working toilets in the school	5.79	10.5	4.71**	5.37	13.29	7.9***
12.	Schools with separate toilets for girls and boys, %	89.0	86.0	-3.0	95.0	100.0	5.0

Notes: Diff. = Difference between urban and rural scores for the indicators; Total number of rural schools is 19, and urban schools is 14; t-test of the differences is significant at **p<0.05; *p<0.1

In terms of teacher quality, as shown in Table 5, the trends were broadly positive. The proportion of qualified and specialised teachers improved notably, as did the average levels of mathematics content knowledge, and the proportion of teachers who had completed level 2 CPD training. However, teachers were more likely to be younger and less experienced in 2018-19 than in 2012-13. A closer look at the urban-rural differences across the 2012-13 and 2018-19 cohorts reveals a similar trend. The proportion of mathematics teachers with a diploma or university degree, teachers who completed level 2 CPD training, and teachers' mathematics content knowledge improved notably for rural schools compared to those in urban settings. For example, the 2012-13 urban-rural gap in teachers' mathematics content knowledge was 44 points (nearly half a standard deviation, in favour of teachers from urban schools). For the 2018-19 cohort, however, the urban-rural gap reversed, with rural teachers scoring 11 points higher. While this is somewhat inexplicable, it is possible that the GEQIP-II related educational reforms on the provision of in-service training benefited rural teachers more than their urban counterparts. Table 5 provides information on the statistical test of significance.

Table 5. Teacher quality indicators related to GEQIP-II reforms, by rural-urban location

		2012-13			2018-19		
		Rural	Urban	Diff.	Rural	Urban	Diff.
1.	Proportion of male mathematics teachers, %	50.0	57.0	7.0	50.0	62.0	12.0
2.	Teachers' age, average	31.39	38.93	7.54**	29.69	33.38	3.70
3.	Teachers' years of teaching experience, average	9.77	15.33	5.56*	4.41	6.2	1.79
4.	Proportion of teachers with a diploma/ university degree/ teacher training qualification, %	72.0	64.0	-8.0	100.0	77.0	-23.0**
5.	Proportion of teachers who completed level 2 CPD training, %	50.0	43.0	-7.0	50.0	69.0	19.0
6.	Proportion of teachers who specialised in mathematics, %	17.0	21.0	4.0	71.0	100.0	29.0**
7.	Teacher experience of teaching Grade 4, average	9.28	18.36	9.08**	4.0	5.92	1.92
8.	Teacher's mathematics content knowledge, average	460.0	504.0	44.0*	520.0	510.0	-10.0

Notes: Diff. = Difference between urban and rural scores for the indicators; total number of teachers in rural schools is 19, and in urban schools is 14; t-test of the differences is significant at ***p<0.001; **p<0.05; *p<0.1

Discussion

The GEQIP-II reforms (2012-2018) in Ethiopia focused on improving quality, equity, and learning outcomes through investment in critical areas of the general education system. Within this context, we provide descriptive statistics relating to scores on mathematics tests administered at the beginning and end of Grade 4 in a purposefully selected sample of 33 schools in six regions of Ethiopia. We found that learning levels declined over the six-year GEQIP-II reforms period, while pupils demonstrated slightly higher learning progress over a school year. There is some evidence of differences in learning progress among pupils across regions and rural-urban localities. In turn, this generates two questions: (1) "Why did mathematics learning levels decline over the six-year period while pupils made relatively higher progress in the 2018-19 school year compared to 2012-13?"; and (2) "Why did progress in mathematics over the school year differ for the two periods between pupils across rural-urban locations, regional states, and family economic backgrounds?"

We note the following: on the one hand, the overall decline in learning levels, the continued gap in learning levels between what are termed established regions (such as Addis Ababa) and emerging regions (e.g., Somali), and between urban and rural pupils appear to raise questions regarding the equity effects of the GEQIP-II reforms. Looking at the learning levels alone, one may conclude that pupils from rural schools, from the poorest socio-economic backgrounds, and from emerging regions may not have benefited equally from the GEQIP-II reforms. That said, there is also evidence of changes in student composition between 2012 and 2019. For example, there is an increase in enrolment in the Somali region among students from disadvantaged backgrounds, which may explain the lowest learning levels for pupils in this region in 2019 at the end of Grade 4. As previously noted, pupils in the Somali region made relatively lower progress over the 2018-19 school year, and performed the lowest both at the start and end of the school year compared to the other regions.

Accordingly, the decline in mathematics scores should not be seen as a failure of the GEQIP-II educational reforms, although our findings do raise questions about the design and scale of reforms that may be needed to improve learning outcomes for all. As shown above, there is little evidence to suggest that school and teacher quality worsened between 2012 and 2019. In fact, there is some evidence that some of the key school and teacher quality indicators improved over the reform period. We also found that the Grade 4 2018-19 cohort (right after the GEQIP-II reforms began) was more likely to have attended pre-school, have lower absence rates, and have fewer episodes of dropout compared to pupils in Grade 4 in 2012-13 (prior to the reforms). These pupil-level improvements are linked to the GEQIP-II reforms in terms of supporting access and student retention. At the same time, there are several indications to suggest that the enrolment of pupils from relatively disadvantaged backgrounds increased between 2012 and 2019. It could be that the influx of more disadvantaged pupils (i.e., those from the poorest economic backgrounds, with caregivers who are illiterate) into the primary education system contributed to the decline in average learning levels. As literate or educated caregivers are more likely to make a greater investment in their children's education (e.g., Rolleston, 2014), the decline in the proportion of literate caregivers for rural pupils may have contributed to the decline in learning levels. Besides, given the expansion in enrolment among the most disadvantaged children in Ethiopia, it is possible that pupils in 2018-19 may have entered Grade 4 under-prepared to learn, compared to those in 2012. Overall, given that in some respects pupils became more disadvantaged over time, the fact that progress in mathematics improved slightly for the 2018-19 RISE cohort and notably so for the RISE rural cohort suggests that the potentially negative effects of home disadvantages were overcome by countervailing factors, including some of the GEQIP-II related improvements in school and teacher quality.

It should also be noted that improvements in key measures of school infrastructure and teacher characteristics associated with the GEQIP-II reforms may take a significant amount of time to produce improvements in learning levels in a system that is expanding rapidly while simultaneously targeting more children from disadvantaged backgrounds. In countries such as Ethiopia, where enrolment among the most disadvantaged is increasing rapidly, the immediately observable effects are reductions in start-of-school year test scores. Educational reforms such as GEQPI-II that target school environments as sites for systematic intervention may well be promising responses to such trends in the long term.

In turn, our findings suggest that the equity dimensions of GEQIP-II may be well-chosen, with pupils from rural schools being able to make significantly higher progress in mathematics over the 2018-19 academic year compared with similar learners prior to the reforms (2012-13). It is possible that the decline in start-of-school year test scores is driven by the effects of expansion into more marginalised segments (e.g., rural areas, deprived economic backgrounds, and emerging regions, in this case). This enrolment profile, combined with improvements in teacher qualifications for schools in rural areas, strong improvements in rural school teachers' mathematics content knowledge, and the notable progress in mathematics over time in favour of pupils from relatively poorer economic backgrounds could instead be viewed as both access and equity successes of GEQIP-II in Ethiopia.

We should note, however, that pupils from relatively advantaged backgrounds who were on target to succeed seem to fail to continue to make comparable learning progress over time. With urban pupils in the 2018-19 cohort, for example, learning progress did not improve, compared to the same group in 2012-13; and pupils from relatively wealthier economic backgrounds made slightly lower progress in mathematics in 2018-19 compared to that made by the same group in 2012-13, prior to the GEQIP-II reforms. Moreover, mathematics content knowledge for teachers of Grade 4, urban pupils in 2018-19 did not improve compared to the same urban group in 2012-13, although the content knowledge

for the 2018-19 rural teachers saw notable improvements. A key question, therefore, is whether some pupils are made ‘worse off’ by expansion, even though a large number are much better off (those who had very limited access previously). Ensuring that more pupils benefit from educational reforms without disadvantaging historically more advantaged pupils (e.g., urban pupils and those from relatively wealthier backgrounds) seems to be a reasonable and just strategy for future education reforms. Our findings appear to suggest that GEQIP-II has fallen slightly short on ‘raising the roof’, and going forward this may be a lesson for Ethiopia and for other countries undertaking reforms with similar aims and within similar contexts. The RISE Ethiopia team continues to engage with comprehensive research based on our unique longitudinal dataset and to produce additional evidence to challenge the interpretation of educational reforms as a deficit. This, and other ongoing studies, are expected to provide answers to one of the main research questions of the RISE Ethiopia research project: What are the key inputs of educational reforms (e.g., the ongoing GEQIP for Equity/GEQIP-E) that improve learning outcomes, particularly for the most disadvantaged children?

References

Aurino, E., James, Z., and Rolleston, C. 2014. Young Lives Ethiopia School Suvey 2012-13. Data Overview Report (Working Paper 134). Oxford: Young Lives.

Hoddinot, J., Iyer, P., Sabates, R., and Woldehanna, T. 2020. Evaluating Large-Scale Education Reforms in Ethiopia. RISE Working Paper Series. 19/034. https://doi.org/10.35489/BSG-RISE-WP_2019/034

Ministry of Education. 2010. Education Sector Development Programme IV (ESDP IV): 2010/11 -2014/15. Program Action Plan.

Ministry of Education. 2015. Education Sector Development Programme V (ESDP V): 2015/16 -2019/20. Program Action Plan. Addis Ababa. Federal Ministry of Education.

Ministry of Education. 2020. Education Statistics Annual Abstract. September 2019-March 2020. file:///C:/Users/dtt24/Downloads/Education Statistics Annual Abstract March 2019 - September 2020.pdf

NEAEA. 2016. Ethiopian Fifth National Learning Assessment of Grades 4 and 8 Pupils. Addis Ababa. National Educational Assessment & Examinations Agency.

Rolleston, C. 2014. Learning profiles and the “skills gap” in four developing countries: A comparative analysis of schooling and skills development. Oxford Review of Education, 40(1), 132–150. https://doi.org/10.1080/03054985.2013.873528

Rolleston, C., James, Z., Pasquier-Doumer, L., and Tam, T. N. T. M. 2013. Making Progress. Report of the Young Lives School Survey in Vietnam. Working Paper 100. Oxford: Young Lives.

Tiruneh, D.T., Hoddinott, J., Rolleston, C., Sabates, R. and Woldehanna, T. 2021. Understanding Achievement in Numeracy Among Primary School Children in Ethiopia: Evidence from RISE Ethiopia Study. RISE Working Paper Series. 21/071. https://doi.org/10.35489/BSG-RISE-WP_2021/071

Tiruneh, D.T., Sabates, R., and Woldehanna, T. 2021. Disadvantaged Schools and Students in Ethiopia: Why is the GEQIP-E Reform Necessary? 2021/026. https://doi. org/10.35489/BSG-RISE-RI_2021/026

UNICEF. 2017. Education sector budget brief - 2016/17. UNICEF for every child Ethiopia. https://www.unicef.org/ethiopia/sites/unicef.org.ethiopia/files/2020-01/National_Education_Budget_Brief_2016_17update.pdf

USAID. 2019. Early Grade Reading Assessment (EGRA) 2018 Endline Report. Addis Ababa. USAID.

World Bank. 2013. General Education Quality Improvement Project II (GEQIP-II). Project Appriasal Document.

World Bank. 2017. General Education Quality Improvement Program for Equity (GEQIP-E): Project Appraisal Document: Report No: 121294-ET. The World Bank.

Dawit T. Tiruneh is a Research Associate at the Faculty of Education, University of Cambridge. In February 2022, Dawit joined the Andrew and Virginia Rudd Research and Professional Practice Programme, where he has worked across a variety of projects. His main research focus is on the promotion of mental wellbeing of children and adolescents in the UK. Designing and evaluating evidence-based interventions aimed at improving educational outcomes of children and young people with mental health issues is among his research interests.

Ricardo Sabates is the Research Management Lead on the RISE Ethiopia team. He is a Professor of Education and International Development at the University of Cambridge and member of the Research for Equitable Access and Learning (REAL) Centre. His research relates primarily to questions of equity in educational access and learning in sub-Saharan Africa and on second chance opportunities for drop-out children particularly in Ethiopia and Ghana. He is currently the co-investigator of a number of research projects within the REAL Centre including secondary education in Africa and has undertaken a number of quantitative projects in education including evaluations of policy reforms.

Caine Rolleston is a member of the RISE Intellectual Leadership Team, a member of the RISE Ethiopia team, and a member of the RISE Vietnam team. He is a Senior Lecturer in Education and International Development at University College London Institute of Education (UCL-IOE). He has worked on education and international development in a range of countries including Ghana, Vietnam, Ethiopia, Peru, India and Sri Lanka, and is currently Senior Education Associate for the Young Lives comparative international study of childhood poverty, based at the University of Oxford. For Young Lives, he leads the development of school surveys and research on school effectiveness. His research interests include issues in the economics of education in developing countries, educational access and equity, privatisation, learning metrics and trajectories, longitudinal studies in education and development, cognitive and non-cognitive skills development and survey design.

John Hoddinott is a researcher on the RISE Ethiopia team. He is the H.E. Babcock Professor of Food & Nutrition Economics and Policy at Cornell University. His research portfolio includes work on the linkages between nutrition and schooling and social protection and schooling. He brings expertise in the design and implementation of large scale quantitative surveys in rural Ethiopia, analysing these data using advanced econometric and impact evaluation methods, undertaking qualitative fieldwork (such as key informant interviews) and presenting these findings to government officials and its donor partners. Between 2004 and 2009, in collaboration with Professor Woldehanna at the Policy Studies Institute, he led the implementation of the Ethiopian Rural Household Survey – the most well used longitudinal survey in Africa and between 2006 and 2015, he led the evaluation of Ethiopia's Productive Safety Net Programme.

Citation:

Tiruneh, D.T., Sabates, R., Rolleston, C.and Hoddinott, J. 2022. Trends in Mathematics Learning in Ethiopia: 2012 - 2019. RISE Insight Series. 2022/045. https://doi.org/10.35489/BSG-RISE-RI_2022/045