

Understanding teachers' contribution to learning outcomes: Implications for teachers' policy in Ethiopia

Moses Oketch^{a,*}, Caine Rolleston^a, Cesar Burga Idrogo^a, Mesele Araya^b

^a University College London, United Kingdom

^b International Growth Centre, United Kingdom

ARTICLE INFO

Keywords:

Teacher value added
Value added modelling
Raising learning outcomes
Ethiopia
GEQIP
Learning gain

ABSTRACT

Estimating teachers' relative contribution to raising learning outcomes at scale is important for informing policy on teachers and teacher development. Much of the research dedicated to assessing the contribution of teachers to their pupils' progress in quantitative terms uses value-added modelling (VAM). Value-added modelling is typically estimated based on students' measured learning gains over a course of a fixed period such as an academic year. Unfortunately, value-added estimates generally describe a black box of mechanisms and do not identify which specific teacher practices and/or interactions with pupils lie behind students' progress. Therefore, it is worth asking whether and to what extent VAM can be used to understand teacher contribution to pupils' progress, and how it can help us to inform teacher policy. In this article, we examine both questions in the context of Ethiopia. We argue that VAM offers important potential, but that it needs careful contextualization to adequately inform teacher policy. While data suited for value-added analysis are rarely available in low- and middle-income contexts, we are able to make use of such data from the Research on Improving Systems of Education (RISE) Ethiopia programme. Ethiopia is a particularly interesting case study. Important reforms to improve both equity and quality of basic education have been implemented, with a strong focus on strengthening teacher training. Nonetheless, learning outcomes have declined in recent years. Our findings provide additional evidence supporting the critics of unconditional VAM approaches.

1. Introduction

Estimating teachers' relative contribution to raising learning outcomes at scale is important for informing policy on teachers and teacher development. Much of the research dedicated to assessing the contribution of teachers to their pupils' progress in quantitative terms uses 'value-added modelling' (VAM). 'Value-added' is typically estimated based on students' measured learning gains over a course of a fixed period such as an academic year.

Unfortunately, simple unconditional value-added estimates generally describe a black box of mechanisms (Oketch et al., 2021) and do not identify which specific teacher practices and/or interactions with pupils lie behind students' progress. This way of analysing teachers' contribution may have important implications in terms of equity as literature has shown that pupils' progress depends on an important set of factors, some of which are beyond teachers' control (Barnes et al., 2018). It also has important limitations for policymaking regarding teachers, since

simple estimation of teachers' value-added does not provide us with information about what should be improved to increase teachers' effectiveness. This raises questions about the utility of VAM for estimating and improving teachers' performance (McCaffrey, Education., 2003). In this article, we examine the usefulness of VAM approach in the context of Ethiopia. We argue that VAM offers important potential to understand teachers' contribution to pupils' outcomes, but that it needs careful contextualization to adequately inform teacher policy.

While data suited for value-added analysis are rarely available in low- and middle-income contexts, we are able to make use of unique data from Research on Improving Systems of Education (RISE) Ethiopia programme. We use data from pupils in primary Grade 4 collected in 2018–19, which includes information on learning outcomes in mathematics both at the beginning and end of the academic year, as well as on pupils' backgrounds, and teachers' characteristics. Ethiopia is a particularly interesting case study. Important reforms to improve both equity and quality of basic education have been implemented in recent years,

* Corresponding author.

E-mail addresses: m.oketch@ucl.ac.uk (M. Oketch), c.rolleston@ucl.ac.uk (C. Rolleston), cesar.idrogo.22@ucl.ac.uk (C.B. Idrogo), mesele.araya@theigc.org (M. Araya).

<https://doi.org/10.1016/j.ijedudev.2025.103425>

Received 9 May 2025; Received in revised form 9 October 2025; Accepted 14 October 2025

Available online 23 October 2025

0738-0593/© 2025 The Author(s). Published by Elsevier Ltd. This is an open access article under the CC BY license (<http://creativecommons.org/licenses/by/4.0/>).

with a strong focus on strengthening teacher training (World Bank, 2008; 2013). Nonetheless, learning outcomes have declined.

Different explanations for this apparent paradox have been provided in the literature. Some of them highlight the increasing number of pupils from disadvantaged backgrounds and the greater challenges they face in the learning process (Iyer et al., 2020; Oketch et al., 2021; Rolleston et al., 2025). Others have noted that, although overall learning outcomes have declined, pupils' progress by the end of the year has actually improved and have linked this modest progress to some aspects that recent educational reform has implemented, such as improved teachers' subject knowledge (Hoddinott et al., 2024). In this article we attempt to add to understanding of which aspects of teachers' development contribute to students' learning progress in Ethiopia. As a low-income country, resources in Ethiopia are very limited and therefore efficient use of them is of utmost importance. Exploring which teachers' characteristics and practices are associated with greater pupils learning progress could, in principle, contribute significantly to improving the design and implementation of future reforms. We reflect on the potential of VAM in this connection.

The rest of this article is structured as follows. Section 2 provides a brief overview of the Ethiopian education context, with a focus on the paradox of declining learning outcomes despite improvements in education quality and teaching delivery. In Section 3, we present a literature review on three key topics relevant to this study: the literature on the strengths and limitations of the VAM framework, the recent literature using VAM in developing countries and its relationship with educational policy, and the literature on the predictors of learning outcomes in the Ethiopian context. In Section 4, we present the methodology we will use to better understand teachers' contribution to the learning process, which is based on a conditional VAM estimation, and we explain why we consider this a more suitable approach than a simple VAM. Section 5 presents the results of our analyses, and Section 6 discusses our findings and their implications for teacher development interventions, in relation to the existing literature.

2. The Ethiopian context: improving quality and declining outcomes

In recent years, Ethiopia has undertaken nationwide comprehensive reform of its educational system by implementing the General Education Quality Improvement Program (GEQIP).

In 2008, the Government of Ethiopia launched GEQIP as a multi-donor initiative aimed at improving both quality and equity in the country's general education system (World Bank, 2008; 2017). This programme initially benefited approximately 15.9 million students and 225,000 teachers across primary and secondary schools (World Bank, 2008). It was implemented in several phases. The first two, GEQIP I (2008–2013) and GEQIP II (2013–2017), focused on improving the quality of educational provision, backed by substantial funding (about US\$417 million in Phase I and US\$550 million in Phase II) (World Bank, 2013; 2017). During these phases, school grants were scaled up and the proportion of teachers with at least a diploma or specialized training increased significantly, as many certificate-level teachers upgraded their qualifications. The provision of continuous professional development (CPD) for teachers also expanded, and the use of information and communication technology (ICT) in teaching was strongly encouraged - for example, secondary schools were provided with ICT facilities under GEQIP II (World Bank, 2008; 2013). The final phase of this reform, GEQIP-E (2018–2023), had an explicit focus on equity and was designed to ensure that pupils from less advantaged backgrounds receive support to achieve learning outcomes more comparable to those of their more privileged peers (World Bank, 2017). It built on the earlier phases while introducing targeted interventions for disadvantaged regions and student groups and was launched with an initial investment of around US\$440 million (World Bank, 2017).

However, despite substantial investments made through GEQIP over

the past decade, learning outcomes in Ethiopia have remained largely unsatisfactory (Iyer et al., 2020; Oketch et al., 2021; Hoddinott et al., 2024). The literature shows that most pupils do not meet curricular expectations for their grade. Results in both reading and mathematics assessments indicate that about 60 % of pupils in Grades 2 and 3 lack foundational reading skills (NEAEA, 2022), and that their learning outcomes have been declining in recent years (USAID, 2019). As Rossiter et al. (2018) point out, this has resulted in a growing proportion of pupils in Ethiopia advancing to the next grade without acquiring the basic math and reading skills expected by the curriculum. Consequently, by the end of primary education, most students are several years behind curricular expectations.

Oketch et al. (2021) show that by 2016–17, about one third of pupils in Grades 4 and 5 had not attained the most basic literacy skills, and more than 80 % of pupils failed to master the most basic numeracy skills. A majority of pupils in Grades 7 and 8, the penultimate and final grades of primary school, were found to be at least three years behind curricular expectations, with pupils from rural areas particularly affected.

The learning crisis affects many countries across the world, and it is particularly severe in low-income countries (Angrist et al., 2021). However, this persistence of poor outcomes is particularly problematic in the Ethiopian context if we consider the intentions of the changes introduced as part of GEQIP, including the measures taken to improve teachers' training and quality of teaching and the incentives and resources aimed at raising teaching standards (World Bank, 2008; 2013.).

Indeed, as Tiruneh et al. (2022) show, since the implementation of the reforms an important set of indicators associated with teachers' quality has improved. For instance, the share of teachers holding a degree, a diploma or specialized teacher training increased from 72 % to 100 % in rural areas, and from 64 % to 77 % in urban areas between 2012/13 and 2018/19. During the same period, the proportion of teachers completing continuous professional development courses (CPD) increased from 43 % to 69 % in urban areas. The share of teachers who taught maths and had specialized in mathematics rose from 17 % to 71 % in rural areas and from 21 % to 100 % in urban areas, linked to an increase in teachers' mathematics knowledge on average.

It seems paradoxical that in a context where several indicators of teaching quality have improved, pupils' learning outcomes have nonetheless declined (Araya et al., 2023; Hoddinott et al., 2024). It is of course possible that these improvements to 'inputs' have yet to feed through into 'outputs' and that these improvements are in part counter-balanced by other trends including rising enrolment among more disadvantaged pupils. A fuller understanding of teachers' contribution to learning outcomes is therefore required to disentangle this apparent paradox, and at the same time to provide valuable insights into which aspects of the GEQIP reforms aimed at improving teacher quality may have worked or which have not.

3. Literature review

3.1. Value-added models and the challenge of estimating teachers' contribution

Much research assessing how teachers contribute to pupils learning progress makes use of a value-added modelling (VAM) approach. The VAM approach emerged in the 1970s (Armor et al., 1976; Murnane, 1975) and has since been used increasingly by academics as well as by policy makers, to assess both school and teacher quality (Hanushek and Rivkin, 2012; Isenberg and Hock, 2010; General Assembly of Pennsylvania, 2002).

VAM consists in aggregating pupils' individual progress at class or school level and using this aggregated progress ('raw' value-added) to model quality of teachers or schools (Chetty et al., 2014; Perry, 2016). Most of the research assessing teachers from a VAM approach has been conducted in developed countries, largely due to availability of relevant data and the emphasis on teacher accountability in these contexts.

While VAMs can be informative about the overall progress of a group, outcomes should be interpreted with caution when used to assess teachers' contribution. In an early critique of the VAM approach, McCaffrey and Education (2003) questioned whether most of literature defending the use of simple VAMs to measure the "causal effect" of teachers on pupils' progress used adequately rigorous methodologies. They pointed out that many studies based their conclusions on informal meta-analysis (see for instance Wright et al., 1997) and arguably omitted important sources of bias in their estimations (see Rowan et al., 2002) or analysed learning gains using tests with different scales (see Hanusek et al., 1998). One consequence of such approaches was an overestimation of the role played by teachers in explaining pupils' progress (McCaffrey and Education, 2003).

More contemporary research has pointed in the same direction, showing that a considerable share of pupils' outcomes depends on conditions which are beyond teachers' control. For instance, research exploring school-level determinants of learning outcomes often points to the availability of pedagogical resources in classroom, the design of the curricula and the size or the composition of the class (Barnes et al., 2018; Gulosino, 2018; Kaffenberger & Pritchett, 2021; Herrmann et al., 2016).

In terms of pupil-level characteristics, Dearden et al. (2011) showed that home background is of major importance. Prior academic achievement is also one of the most important factors shaping pupils' learning outcomes. It captures a considerable part of the effects of socioeconomic status in previous periods, including parental educational background and the availability of resources at home and the quality of education received previously (Chetty et al., 2014). It also captures individual effort and ability attributes to the extent that these influence prior performance. 'Sorting' or selection of pupils into schools may also play a role as high achievers generally come from more privileged households, with more resources to pay for a higher quality education or may live in areas with better educational opportunities. This may contribute to a potentially misleading initial appearance that these schools or their teachers are more 'effective' than others (McCaffrey and Education, 2003). Affinity between pupils and teachers in terms of gender or ethnic identity has also been found to be associated with better school performance (Dee, 2005).

Another debate regarding the suitability of VAMs estimates concerns the variability of these estimates across time (McCaffrey, Education., 2003). Early research showed that teachers' effectiveness varies with experience (Shkolnik et al., 2002). In a more recent study, Stacy et al. (2018) found that measures of teachers' value added are weakly correlated one year to the other, suggesting that time-varying confounders may shape to a considerable extent the outcomes of their pupils. This is in line with literature which shows that changes in school or class composition are associated with variation in learning outcomes across time (Iyer et al., 2020).

As unconditional VAMs estimates are questioned for both their accuracy and stability, McCaffrey, Education., (2003) highlighted that they should not be considered as recovering causal effects of teachers on pupils' outcomes unless all variables that can introduce bias in the estimation of teachers' contribution are considered. A more radical view is that of Meyer (1997) who argues that it is almost impossible to isolate teacher effects from other contextual factors, and therefore that the notion of estimating the "causal effect" of teachers (their direct contribution to pupils' outcomes) should be abandoned.

However, more nuanced positions hold that VAMs can still be useful for diagnostic purposes, provided they account for the various predictors of pupils' outcomes in order to properly identify the teacher's contribution (Darling-Hammond, 2015). In other words, they should adopt a 'conditional' approach. Conditional VAMs are nonetheless not a perfect solution. It may be argued that this approach can lead to inappropriately lower expectations for disadvantaged pupils, which may be self-reinforcing (Oketch, 2021). However, if used for improving our understanding of the challenges that teachers may face in teaching delivery, VAM can be a potentially useful tool to inform teacher-policy.

3.2. VAM approach in developing countries

In recent years, increasing attention has been given to the VAM approach in developing countries. In Chile, for instance, significant research has been conducted to diagnose the challenges faced by the Chilean educational system in achieving equity. Findings suggest substantial variation in teacher effectiveness in primary education, depending on both pupil-level factors (such as pupils' background) and school-level factors (such as school fees) (Urrea, 2018). Variation in teacher effectiveness has also been observed in secondary education, where a greater proportion of 'effective' teachers work predominantly in high-SES contexts (Torres, 2018). Evidence of an uneven distribution of teacher effectiveness across educational institutions has also been found in Argentina, where the 'best' schools were found to be almost exclusively attended by the wealthiest students (Cervini, 2006).

In other Latin American countries, such as Peru, research using a VAM framework is more recent. Eigbiremolen et al. (2020), for example, used VAM to assess school-level contributions to pupils' learning outcomes, finding important inter-school variation in effectiveness, particularly between private and public institutions, although a considerable share of this variation was linked to differences in pupils' backgrounds. While research on teachers' contribution to value-added remains scarce, a recent article by Bertoni et al. (2024) provides valuable insights. After estimating teachers' value-added using a conditional approach, the authors assessed the efficacy of recruitment instruments as screening devices, finding that teachers' curriculum content knowledge, as well as their performance in standardized tests, were strongly associated with effectiveness. These findings contrast with those of Cruz-Aguayo et al. (2017) in Ecuador, where no evidence was found of a link between teachers' performance in standardized assessments and students' learning outcomes.

In China, research using VAM has been conducted to evaluate school- and teacher-level contributions to pupils' progress. In terms of inter-school variation, the urban-rural gap was identified as a major factor, while at the teacher level, greater time spent in CPD training and participation in both formal and informal professional development activities were associated with higher pupil progress (Thomas, 2020).

Attempts to move beyond the simple estimation of teachers' value-added and to identify which teacher characteristics are associated with greater effectiveness have also been made in low-income countries, with mixed results. In Pakistan, for instance, Bau and Das (2020) found a strong association between teachers' content knowledge and effectiveness. In contrast, in Northern Uganda, Buhl-Wiggers et al. (2022) did not find any significant relationship between observable teacher characteristics and effectiveness as measured through VAM. In Ethiopia, using data from the Young Lives study, Oketch et al. (2020); (2021) documented considerable variation in teacher effectiveness depending on pupils' characteristics, such as household income. This provides evidence that even equally effective teachers may produce very different learning gains depending on classroom composition.

The growing body of research using a VAM framework highlights the potential of this approach to inform education policy and to help address critical challenges facing educational systems in developing countries. In Chile and Argentina, for instance, research has suggested that without equity-oriented policies to substantially improve the quality of education for disadvantaged groups, learning gaps will continue to widen over time (Torres, 2018; Cervini, 2006). In Peru, VAM-based estimates of teacher effectiveness have been used to assess which of the hiring criteria introduced after the 2012 educational reform help achieve the objective of recruiting the 'best' teachers (Bertoni et al., 2024). In Pakistan, evidence of differential teacher effectiveness has been employed to understand teachers' attitudes towards reforms such as the introduction of performance-based pay (Brown & Andrabi, 2023), with findings generally aligned with evidence from other developing (Munoz Chereau et al., 2020) and developed countries (Scheerens et al., 2003).

In Northern Uganda and China, the identification of differential

teacher effects has been used to evaluate which aspects of recent educational reforms, particularly in teacher training, are associated with greater pupil progress (Buhl-Wiggers et al., 2022; Thomas, 2020). Finally, the evidence of differential teacher effectiveness found by Oketch et al. (2021) in Ethiopia provides a starting point to explore which teacher characteristics or pedagogical practices are more effective in improving pupils' learning outcomes.

3.3. Predictors of pupils' performance in the Ethiopian context

In recent years, extensive research in Ethiopia has provided valuable insights into the factors influencing pupils' learning outcomes. Data from the Fourth National Learning Assessment (NLA) from 2012 for Grades 4 and 8 indicated that learning outcomes varied according to a wide range of individual characteristics. Gender, for instance, played an important role, with boys outperforming girls in many subjects. Language of instruction also matters, with pupils taught in mother tongue performing better than those taught in another language. Age was negatively associated with learning outcomes during this assessment. At the household level, family size and parental educational attainment were both significantly associated with greater pupil achievement. Finally, school-level variables such as commuting time were negatively associated with pupils' progress (NEAEA, 2013).

The Fifth NLA wave, conducted in 2015, reported learning outcomes for pupils from the same grades, and expanded the analysis of learning outcomes predictors. At the pupil level, access to additional educational resources was positively associated with learning outcomes, while region and urban-rural location were also important predictors with urban pupils outperforming their rural peers. The time spent consuming radio, or television seemed to play an important role predicting pupils' outcomes with a greater consumption associated with lower learning outcomes. At household level, living with both parents, access to basic services such as electricity, and higher socioeconomic status were all associated with better performance (NEAEA, 2016). Research focusing on predictors of literacy achievement found similar results, but also highlighted the importance of pupils' familiarity with textbooks, the availability of educational resources, and teachers' characteristics (USAID, 2019).

Longitudinal studies provide further evidence. Tesfay (2012) used various waves of Young Lives data from 2002 to 2015 to analyse the predictors of pupils' learning outcomes in Ethiopia at age 15. She found that high early achievement in maths and in cognitive development tests strongly predicted future learning outcomes. Additional characteristics associated with greater outcomes included a lower time spent at household or productive work, and a lower number of younger siblings. These variables were also found to be positively correlated with other educational outcomes, such as completing primary school (Woldehanna and Hagos, 2015).

Tiruneh et al. (2021) used longitudinal data collected by RISE Ethiopia in 2018/19 to conduct a comprehensive analysis of the predictors of Grade 4 pupils' math scores including at the same time individual, household and school-level variables. They fitted three different types of regression models: first, to identify predictors of baseline scores, then to identify the factors shaping endline scores, and finally to explain endline scores conditional on prior achievement. According to their findings, individual-level characteristics influencing pupils' progress included prior academic achievement, which was positively associated with greater progress by the end of the year; pupil age, with older pupils generally outperforming their younger peers; and the amount of time spent studying outside of school, which was also positively linked to higher learning outcomes. When controlling for individual characteristics, no household-level variables were found to be associated with greater learning outcomes. School-level variables such as a greater availability of educational resources or being located in an urban area were significantly associated with greater learning outcomes. Finally, this research also explored which teachers' characteristics were

associated with greater pupils' progress finding that teacher knowledge of the subject (in this case, maths) was an important predictor, a finding that was also identified by Hoddinott et al. (2024).

Focusing on time-varying predictors of pupils' outcomes, Iyer et al. (2020) found that changes in class composition in terms of pupils' backgrounds across time was an important factor explaining variation of learning outcomes. According to the authors, the decline in learning outcomes in Ethiopia could be explained to an important extent by the increasing access for students with disadvantaged backgrounds and the challenges faced by them and their teachers. These findings are in the same line that the findings from Woldehanna et al. (2016) who showed a considerable disparity and declining in learning outcomes between older and younger cohorts in Ethiopia.

Recently, Ethiopian education has been affected by various external shocks such as armed conflict, the COVID pandemic and environmental disasters (Tiruneh, 2020; Yorke, et al., 2021). Some research has already been conducted to assess how these shocks may have affected learning outcomes. For instance, Sabates et al. (2024) found that, after COVID, pupils were reported as having more disabilities and showing a decline in their learning outcomes. Similarly, Araya et al. (2022) found that pupils returning to school after 'lockdown' demonstrated lower academic skills than the previous cohort. International organisations have warned of an alarming situation: by 2024, drop-out had risen to the extent that at least 8 million children were out of school (UNESCO, 2024; UNICEF, 2024).

3.3.1. Data and methodological approach

Evidence from various contexts, including Ethiopia, shows that pupils' learning outcomes are shaped by a range of individual, household, and school factors. Therefore, an adequate assessment of teachers' contribution should employ a conditional VAM approach that accounts as much as possible the range of influences on pupils achievement, to address potential sources of bias when estimating the contribution of schools and teachers (Darling-Hammond, 2015).

The school survey dataset collected by RISE Ethiopia in 2018–19 can be used for this purpose. This is longitudinal dataset including information from Grade 1 and Grade 4 children, parents/caregivers, teachers and school leaders, covering the regions of Addis Ababa, Amhara, Benishangul Gumuz, Oromia, SNNP, Somali, and Tigray. For a detailed description of the sample, see Araya et al. (2023) and Hoddinott et al. (2019). Information collected from pupils includes tests of academic achievement at the start and the end of the year, as well as a rich set of household and individual background characteristics. Information about teachers includes their sociodemographic characteristics as well as important indicators associated with teacher quality such as experience, educational attainment and specialization, among others. Access to fully anonymised data was provided by RISE Ethiopia, a partnership co-ordinated by the Ethiopian Policy Studies Institute and the University of Cambridge's Research for Equitable Access and Learning Centre (REAL). The RISE Ethiopia partnership gained required ethical approvals for the surveys.

We can use the survey data to disentangle which characteristics of teachers are associated with better learning outcomes. For the purposes of this article, we focus on Grade 4 pupils, and we analyse their performance in mathematics as this data is more directly comparable across Ethiopia than data on literacy. Data from Grade 4 pupils were collected from a sample of 166 schools. In each school, 28 pupils were randomly selected from up to two classes. The average class size in the sampled schools was 53 students, with some schools showing very large classes of more than 100 students. From the original sample of 4144 students assessed at the beginning of the year, a final sample of 3350 Grade 4 pupils remained by the end of the year. We have used analysis of full-cases only, where both learning outcomes and all relevant background characteristics are non-missing and have not used imputation for missing data.

Our conditional value-added approach proceeds as follows. First, we

estimate the expected end of year learning outcome in maths for each pupil using a range of predictors based on the literature and given the availability of data. We include as the main individual-level predictors pupils' prior achievement (at the beginning of the year) in maths, gender, using female as category of reference, and age, grouped in three categories. As pupils in Grade 4 are expected to be between 10 and 11 years old, we have labelled these students as being at the "expected age", pupils less than 10 years old are labelled as "below expected age" and pupils over 11 years old "above expected age", with the "expected age" being used as category of reference.

Household-level variables included are the wealth level and the region where the household is located. Wealth level is a categorization into quintiles of the wealth index variable, which is a 0–1 scaled variable calculated using principal components analysis of households' assets. Quintiles are labelled low, mid-low, middle, mid-high, and high. We use wealthiest households as group of reference. In the case of the region, we used Somali as reference. Finally, we include the size of the class in terms of the number of pupils. Maths test scores at baseline and endline are calibrated on a common scale using item-response methods with a mean at baseline of 500 and a standard deviation of 100 across the pooled sample (baseline and endline).

Pupils' expected outcomes are estimated using a simple OLS model regressing pupils learning outcomes at the end of the year on the set of predictors mentioned. The model used is presented below¹:

$$\text{math_el}_i = \alpha + \beta * \text{math_bl}_i + \gamma X_i + e_i \quad (\text{I})$$

With math_el_i being individual learning outcome for each pupil i , math_bl_i is pupils' baseline score in math, X_i being a set of covariates indicating pupils' circumstances, and e_i is the stochastic error term, hence the estimated learning outcome for each pupil would be defined by:

$$\hat{O}_i = \alpha + \beta * \text{math_bl}_i + \gamma X_i \quad (\text{II})$$

Where $\hat{O}_i$ is the predicted learning outcome for pupil i .

Then we aggregate expected pupil learning outcomes ($\hat{O}_i$), estimated with Eq. (II) to obtain the aggregated expected learning outcomes of the class at teachers level $\hat{O}_j$, with subscript j indicating the teacher. Next, we subtract expected learning outcomes by the end of the year at teacher level from the observed learning outcomes, O_j , obtaining our conditional estimation of teachers' value-added.

Finally, we group teachers into two categories based on this conditional estimation of their value-added: if $O_j - \hat{O}_j > 0$, in other words, if observed aggregated pupils' achievement for a teacher is greater than its expected value, then we consider this teacher 'produces' value-added above the mean level; conversely, if $O_j - \hat{O}_j < 0$, if observed aggregated pupils' achievement for a teacher is below its expected value, then the teacher does not produce 'value-added' on our definition. By grouping them into these two categories, we can identify which teachers are linked with a better than average improvement of their pupils' achievement, accounting for the circumstances they face. Furthermore, as we divide teachers into two simple groups (producing/not producing value-added) aspects such as the number of pupils, do not affect our outcomes.

Finally, we use descriptive statistics to see how these two types of teachers are distributed by region, as well as to explore which teachers' characteristics are associated with greater pupils learning outcomes, how they are related to GEQIP reforms, and how this can help us to inform further policy. Variables included to measure teachers' experience are the total years of experience as teacher, the years of experience at that school, and the years of experience teaching math. Variables about teachers training are their educational attainment, the

specialization in the subject and the training received at school. We have also considered to add some variables to see whether gender and language affinity between teachers and pupils are associated with better outcomes. Gender affinity is defined as whether the teacher and pupil share the same gender, while language affinity refers to whether they share the same mother tongue.

A comparison between the outcomes from our conditional approach and those from an unconditional VAM, based only in differences between pupils' endline and baseline scores, was included in the Appendix section to illustrate how the two approaches differ.

4. Results

Table 1 presents descriptive statistics on pupils' learning outcomes and how they vary depending on the covariates we will include in our model. It also includes a column indicating the statistical significance from a difference in means test between each category of the variables and the category of reference.

Pupils' learning outcomes in Somali region are well below other regions, with differences in outcomes by the end of the year being statistically significant in all cases. The greatest gap is observed between Somali and Addis Ababa (about 160 points), followed by Amhara and Tigray (about 100 points) and Oromiya (about 77 points). Pupils' learning outcomes also differ by level of wealth, with the wealthiest pupils having the greatest scores by the end of the year. Differences in scores between pupils from the first three income quintiles are not large. The age of pupils is also associated with their learning outcomes. Differences by gender are small, but statistically significant, with boys outperforming girls by about 15 points on average. We can also see that older pupils score on average more than pupils from the expected age, while no differences are observed between this group and younger pupils. Regarding the size of the class, we see that pupils in smaller classes, on average, perform better than pupils in larger classes and that these differences are statistically significant, which may indicate that it may be 'easier' to monitor and attend to the learning needs for a small group of pupils.

Fig. 1 shows the relationship between math scores obtained at the end and the start of Grade 4. Overall, we see that pupils that score well at

Table 1
Average learning outcomes by region, individual and classroom characteristics.

		Average endline score	Diff in means signif.	N
Region	Addis Ababa	611.03	***	530
	Amhara	551.29	***	627
	Benishangul	485.30	***	468
	Gumuz			
	Oromiya	527.02	***	1086
	SNNP	494.51	***	569
	Somali	450.35	Ref.	342
	Tigray	550.37	***	522
Wealth Quintile	Low	511.15	***	822
	Mid-low	518.79	***	833
	Middle	509.53	***	803
	Mid-high	535.62	***	848
	High	569.99	Ref.	838
Gender	Female	522.80	Ref.	2006
	Male	537.47	***	2132
Age group	Less than 10 years-old	532.95		432
	From 10–11 years- old	525.35	Ref.	2122
	More than 11 years-old	536.81	*	1578
Size	50 pupils or less	545.80	Ref.	2110
	More than 50 pupils	512.84	***	2034

***p < 0.001; **p < 0.01; *p < 0.05; †p < 0.1

¹ We used STATA for all statistical analyses.

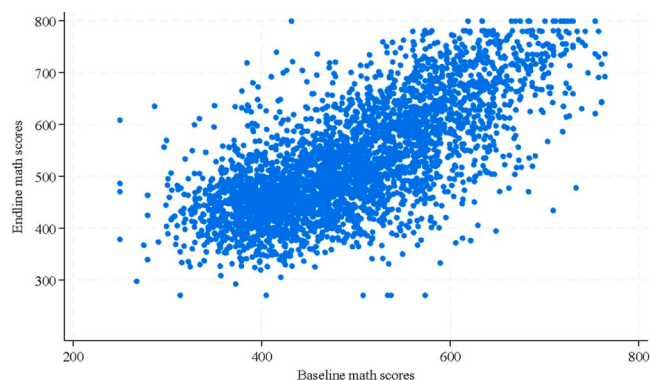


Fig. 1. Baseline scores vs Endline scores.

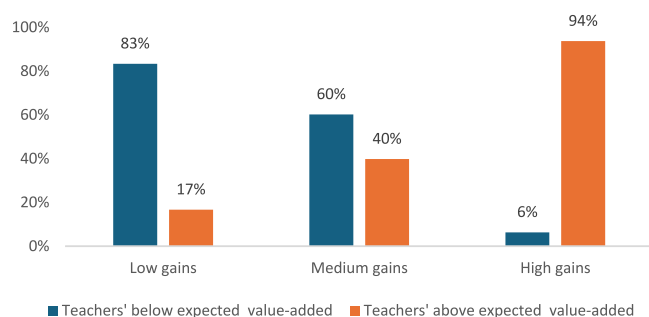


Fig. 2. Teachers by unconditional VAM vs conditional VAM.

baseline also scored well by the end of the year; however, we do not see progress among high achievers. Indeed, the highest score both, at baseline and endline are around 800, which indicates the presence of a 'ceiling effect'. Among low achievers the progress is stronger with many pupils making substantial improvements by the end of the year.

Overall, our descriptive analysis confirms that variation in learning outcomes is considerably shaped by individual, household and school-level variables, supporting the idea that accounting for all these factors is necessary to adequately assess teachers' contribution to pupils' progress.

Table 2 presents the results of the OLS model used to estimate expected learning outcomes of pupils. Prior achievement is, not unexpectedly, the most notable predictor of learning outcomes explaining about 49 % of the variability of the scores by the end of the year (model 1). Once we account for all other individual characteristics, the variance explained increases modestly to 54 % (model 2). Adding the class size does not change the predictive power of the model significantly (model 3).

The final model (model 4) includes only the variables that show statistically significant associations with learning outcomes at the end of the year. Addis Ababa is the highest performing region, with learning outcomes on average 82 points greater than those of Somali. The expected outcomes of pupils in Tigray, Amhara and Oromiya are also considerably greater than in Somali. We also see that, pupils at the middle of the wealth distribution are the ones that perform least well, compared to pupils at the bottom, but particularly to those at the top of the wealth distribution, who remains the best achievers. On the other hand, differences in learning outcomes by age remain similar to those presented in bivariate descriptive analysis from Table 1. The size of the class also plays a similar role with an increase in the size of the classroom by one pupil associated with an average decline in learning outcomes of 15 points, supporting the idea that smaller groups may help teachers to better support pupil. Finally, the relationship between prior achievement and learning outcomes is clear: high baseline achievers are the ones with the greatest performance by the end of the year. This may

be explained by prior knowledge of high achievers which makes it easy for them to learn the topics of the course, but also by uncontrolled differences in social background (parental education, school resources at home, etc).

We now group teachers by whether they produce value-added or not, based on the aggregated expected outcomes of pupils for each teacher (using model 4) conditioned on the characteristics included in our regression model. In Table 3 we see the distribution of teachers producing value-added by region.

Contrary to what might be expected, we see that in Somali we find the greatest share of teachers with a positive value-added. In contrast, in regions like Amhara and Tigray, where higher learning outcomes are observed, the share of teachers whose pupils reach at least the expected learning outcomes is considerably lower.

This apparently surprising outcome can be explained by the conditional approach used. In fact, as we can see in Figure A1 (see Appendix section), the identification of effective teachers differs when we move from an unconditional to a conditional approach with a non-negligible share of teachers (17 %) grouped as low gains when using an unconditional VAM producing value-added in the conditional approach. This proportion is considerably higher across medium-gain teachers (40 %). Thus, when we see the differences in the classification by region, we found that in Somali half of low-gain teachers shows a positive value conditioned to pupils' characteristics (see Appendix A2). Again, this is of major importance from an equity perspective given the huge proportion of teachers in Somali that are grouped as low-gains from a simple VAM framework (see Appendix A1). According to our findings, teachers in Somali would be, overall, making their pupils progress besides the circumstances they have to face.

Finally, let's explore how this conditional approach helps us to better understand whether the changes introduced by GEQIP reform are associated with better learning outcomes. Table 4 shows us the distribution of different variables associated with teachers' quality and training for the two groups of teachers.

We clearly see that teachers who show a positive value-added are the most experienced ones, both in general terms (two years more experienced on average) and in the specific course, in this case, maths (1.8 years more experienced on average). Second, we observe that these teachers are more likely to be specialized in the course they are teaching (79 % vs 70 % across teachers below their expected value-added). Differences by educational attainment are somehow important, although not too much. Having a diploma does not play an important role differentiating teachers' performance and having a degree, while appears to be more associated to a better teachers' performance, does not seem to be a key aspect either. The reason for this may be that our conditional approach accounts for characteristics associated with teachers' educational attainment such as the region where children study (see Appendix A4). Differences are very small by gender affinity, while language affinity appears to be more associated with lower learning outcomes for pupils.

In this regard, it is interesting to note that the association between language affinity and teachers' value-added differ when we use an unconditional VAM (see Appendix A5). This may appear puzzling and counterintuitive; however we should keep in mind that most of teachers labelled as high-gains (effective) from an unconditional VAM are located in regions like Oromiya, Amhara and Tigray, which have also the greatest language affinity (see appendix A3). The fact that association changes once we control for the region in our conditional model may be related to targeting teachers sharing the same language to low-performing pupils precisely with the aim of improving their outcomes. A similar interpretation can be given to the association between greater hours of training and low performing outcomes. As we can see, teachers whose pupils' achievement is below their estimated value receive more training than the best teachers, which may indicate that schools provide an additional support to teachers facing greater challenges. All these school policies would be aligned with GEQIP equity aims.

Table 2

Regression model Dependent variable: Math score by the end of the year.

		M1	M2	M3
Intercept		140.86***	115.96***	124.98***
Prior achievement	Baseline math score	0.79***	0.72***	0.72***
Region	Ref. Somali		-	-
	Addis Ababa		83.64***	82.10***
	Amhara		74.85***	73.77***
	Benishangul Gumuz		49.27***	48.78***
	Oromiya		71.53***	73.14***
	SNNP		43.02***	43.31***
	Tigray		76.31***	74.45***
Wealth quintile	Ref. High		-	-
	Low		-9.88*	-9.85*
	Mid-low		-12.14**	-12.59**
	Middle		-15.13***	-15.56***
	Mid-high		-12.22**	-12.04**
Gender	Ref. Female		-	-
	Male		3.84	3.78
Age	Ref. 10–11 years		-	-
	Less than 10 years-old		2.35	2.26
	More than 11 years-old		11.31***	11.40***
Class size				-0.15*
	<i>N</i>	3359	3350	3350
	<i>Adj. R2</i>	0.49	0.54	0.54

		Model 1	Model 2	Model 3	Model 4
Intercept		140.86***	115.96***	124.98***	126.21***
Prior achievement	Baseline math score	0.79***	0.72***	0.72***	0.72***
Region	Ref. Somali		-	-	-
	Addis Ababa		83.64***	82.10***	81.93***
	Amhara		74.85***	73.77***	73.70***
	Benishangul Gumuz		49.27***	48.78***	48.85***
	Oromiya		71.53***	73.14***	73.16***
	SNNP		43.02***	43.31***	43.17***
	Tigray		76.31***	74.45***	74.20***
Wealth quintile	Ref. High		-	-	-
	Low		-9.88*	-9.85*	-9.75*
	Mid-low		-12.14**	-12.59**	-12.64**
	Middle		-15.13***	-15.56***	-15.65***
	Mid-high		-12.22**	-12.04**	-12.1**
Gender	Ref. Female		-	-	-
	Male		3.84	3.78	
Age	Ref. 10–11 years		-	-	
	Less than 10 years-old		2.35	2.26	2.15
	More than 11 years-old		11.31***	11.40***	11.41***
Class size				-0.15*	-0.15*
	<i>N</i>	3359	3350	3350	3350
	<i>Adj. R2</i>	0.49	0.54	0.54	0.54

Table 3

Share of teachers producing value-added by region.

Total	Addis Ababa	Amhara	Benishangul Gumuz	Oromiya	SNNP	Somali	Tigray
49.4 %	51.4 %	44.7 %	43.5 %	50.8 %	48.6 %	60.0 %	47.2 %

5. Discussion and conclusion

In this article we have explored the potential of VAM to assess understand the contribution of teachers to pupils' learning outcomes as well as to inform policy. The outcomes of our conditional VAM are generally aligned with current literature from Ethiopia stating that pupils' learning outcomes is considerably shaped by individual, household and school-level variables (Tiruneh et al., 2022; Rolleston et al., 2025), although some variables like wealth or gender, which are generally associated with learning outcomes (see NEAEA, 2013, 2016), do not appear to play a significant role from our findings.

One of the most important findings arising from our conditional approach is the identification of a greater proportion of teachers showing a positive value-added in a deprived region like Somali. In fact, a rapid interpretation of teachers value-added based only on final learning gains could lead us to consider that this region has the least

effective teachers (see Appendix A1). However, when we consider the conditions teachers in Somali must face, the picture is more nuanced. In fact, it seems that in Somali, a deprived region where pupils, and schools face several difficulties, there is a greater proportion of teachers whose pupils, overall, make progress. This finding contrasts with most of literature indicating that effective teachers are mostly located in wealthiest areas (Cervini, 2006; Torres, 2018; Urrea, 2018) and could perhaps be explained by two reasons. On the one hand, the recent focus of GEQIP on building a more equitable educational system which means prioritizing the most deprived regions; and on the other hand, the existence of ceiling effects making high-performing students, mostly in Addis Ababa, Amhara or Tigray, hard to move forward in comparison with low-performing students – generally in emerging regions – which have much more space to progress.

This does not mean that we should be satisfied with current learning outcomes in Somali. They are considerably low and, in many cases,

Table 4

Characteristics of teachers producing/not producing value-added.

			Not producing value-added	Producing value-added
Teachers' experience	Years of experience	General experience	8.9	11.2
		Experience at school	5.1	5.5
		Experience at course	4.6	6.4
Teachers' training	Level of education attained	Not trained	3 %	2 %
		Certificate	11 %	6 %
		Diploma	79 %	80 %
		University degree	7 %	12 %
	Specialized in the course	No	30 %	21 %
		Yes	70 %	79 %
	Received training	None	20 %	34 %
		Less than 10 h	12 %	12 %
		11–30 h	11 %	6 %
		31–60 h	46 %	39 %
		61–100 h	6 %	5 %
		More than 100 h	6 %	4 %
Affinity	Gender affinity	No	50 %	49 %
		Yes	50 %	51 %
	Language affinity	No	11 %	8 %
		Yes	89 %	82 %
		N	130	127

deteriorating. But it should be pointed out that poor learning outcomes in this region appears to be explained by factors beyond teachers' "effectiveness", and, in consequence, it should be unfair to hold only responsible to teachers for low pupils' performance. A misjudgement of teachers' contribution does not only have implications in terms of fairness or equity. It may also lead to a misallocation of resources when designing interventions (Giffin et al., 2009; Jearld, 2009). This is of particular concern for the Ethiopian context given it is a low-income country and is not very likely that massive expenditures such as GEQIP reform would be repeated.

In terms of teachers' characteristics and practices associated with better pupils outcomes, the conditional VAM also allows us to have a better idea of which reforms at teachers' level introduced by GEQIP may or may not have been effective for improving learning outcomes, and how this relates to other experiences in developing countries.

For instance, GEQIP reform has encouraged hiring teachers with at least a diploma. Our findings suggests that the requirement of formal qualifications introduced in GEQIP I and II (World Bank, 2008;2013) may not be as important as it sound to guarantee the improvement of pupils' achievement. This is an important finding as many educational reforms in developing countries over the last years have focused on educational credentials (for the case of Peru see Bertoni et al., 2024). In the Ethiopian case, like in Northern Uganda (Buhl-Wiggers et al., 2022) the association between credentials and performance seems to be modest.

In contrast, teachers' experience appears to play an important role to improving learning outcomes. This differs from the picture found in Northern Uganda where teachers' experience was insufficiently associated with higher teachers' value-added (Buhl-Wiggers et al., 2022). This is a particularly interesting point as Tiruneh et al. (2022) have shown that the average years of experience of teachers after the introduction of GEQIP reforms declined in Ethiopia. This was, of course, an inevitable result as an increasing number of pupils demanded more and more teachers not all of which could be highly experienced (Filmer, 2023). However, it is important to keep in mind that, if experience is a major determinant of teachers' performance, further policy on teachers' development should account for this overall reduction of total experience². So far, GEQIP has constantly provided training for teachers,

which seems to have been taken specially by those teachers who are struggling more to meet expectations, however as the expansion of education has been greater in most disadvantaged regions (Rolleston et al., 2025), challenges faced by pupils and teachers require more comprehensive interventions.

A way to counteract the negative effects of low experienced teachers and the increasing difficulties of pupils in learning may be to assure that teachers are specialized in the subject and have a greater domain of it. Indeed, our analysis show that pupils from teachers with a greater domain in maths perform better than their peers from non-specialized teachers, a finding that is also supported by recent literature in Ethiopia (Hoddinott et al., 2024) but also in other developing countries (Torres, 2018; Bertoni et al., 2024). Improving teachers' specialization and knowledge of mathematics has been a strong point of GEQIP reform. As Tiruneh et al. (2022) pointed out, maths knowledge of teachers increased considerably between 2012/13 and 2018/19. It is important then targeted policies for those teachers who are still below an adequate domain maths.

While more detailed research is necessary to have a clearer picture of the extent to which GEQIP reforms to improve teachers' quality have been effective, the use of a conditional VAM can help us to make sense of current data, and to better identify the areas that should be addressed by further teachers' policy.

CRediT authorship contribution statement

Moses Oketch: Writing – review & editing, Supervision, Investigation, Funding acquisition, Formal analysis, Conceptualization. **Caine Rolleston:** Writing – review & editing, Validation, Methodology, Investigation, Funding acquisition, Formal analysis. **Cesar Burga Idrogo:** Writing – original draft, Methodology, Formal analysis. **Mesele Araya:** Validation, Formal analysis, Data curation.

Declaration of Generative AI and AI-assisted technologies in the writing process

There has not been use of AI to write this article.

² <https://documents1.worldbank.org/curated/en/099032824125516424/text/P5010171dfde1d0f1986d11db26894d78a.txt>

Appendix

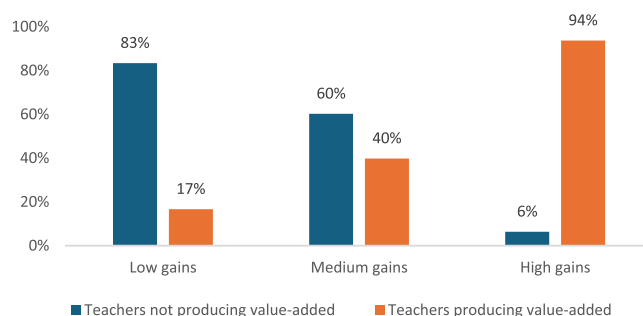


Figure A1. Teachers grouped from unconditional VAM vs teachers grouped from conditional VAM*. *Low gains: bottom 30 % of teachers; middle gains: middle 40 % of teachers; high gains: top 30 % of teachers

A1. Distribution of teachers' unconditional value-added groups by region

	Low gains	Medium gains	High gains	N
Addis Ababa	24 %	41 %	35 %	37
Amhara	20 %	53 %	28 %	40
Benishangul Gumuz	35 %	39 %	26 %	23
Oromiya	12 %	47 %	41 %	59
SNNP	46 %	35 %	19 %	37
Somali	80 %	12 %	8 %	25
Tigray	22 %	39 %	39 %	36
N	77	103	77	257

*Low gains: bottom 30 % of teachers; middle gains: middle 40 % of teachers; high gains: top 30 % of teachers

A2: Share of teachers' producing value-added by region and unconditional value-added grouping

	Addis Ababa	Amhara	Benishangul Gumuz	Oromiya	SNNP	Somali	Tigray
Low gains	0.0 %	0.0 %	25.0 %	0.0 %	5.9 %	50.0 %	0.0 %
Medium gains	40.0 %	33.3 %	33.3 %	28.6 %	76.9 %	100.0 %	28.6 %
High gains	100.0 %	100.0 %	83.3 %	91.7 %	100.0 %	100.0 %	92.9 %

*Low gains: bottom 30 % of teachers; middle gains: middle 40 % of teachers; high gains: top 30 % of teachers

A3: Share of pupils sharing same language as their teachers by region

	% of pupils sharing same language as their teachers
Addis Ababa	83 %
Amhara	97 %
Benishangul Gumuz	58 %
Oromiya	93 %
SNNP	75 %
Somali	80 %
Tigray	95 %
Average	85 %

A4. Teachers' educational attainment by region

Table A1

Distribution of unconditional VAMs teachers' groups by region

	Not trained	Certificate	Diploma	University degree
Addis Ababa	5 %	0 %	78 %	16 %
Amhara	2 %	18 %	80 %	0 %
Benishangul Gumuz	0 %	8 %	88 %	4 %
Oromiya	2 %	5 %	73 %	20 %
SNNP	6 %	0 %	94 %	0 %
Somali	0 %	32 %	64 %	4 %
Tigray	0 %	5 %	85 %	10 %
Total	2 %	9 %	80 %	9 %

A5. Characteristics of teachers by effectiveness group using and unconditional VAM*

			Low gains	Middle gains	High gains
Teachers' experience	Years of experience	General experience	9.1	10.9	9.9
Teachers' training	Years of experience	Experience at school	6.5	4.7	4.8
		Experience at course	6.3	5.5	4.7
		Not trained	3 %	3 %	1 %
		Certificate	13 %	9 %	4 %
		Diploma	80 %	78 %	81 %
	Level of education attained	University degree	4 %	10 %	14 %
		Specialized in the course	No	31 %	22 %
		Yes	69 %	78 %	74 %
		Received training	None	20 %	20 %
		Less than 10 h	15 %	11 %	10 %
		11–30 h	11 %	9 %	6 %
		31–60 h	37 %	51 %	36 %
		61–100 h	11 %	3 %	3 %
		More than 100 h	7 %	5 %	3 %
Affinity	Gender affinity	No	49 %	50 %	50 %
		Yes	51 %	50 %	50 %
		Language affinity	No	16 %	12 %
	Language affinity	Yes	84 %	84 %	88 %
		N	77	103	77

*Low gains: bottom 30 % of teachers; middle gains: middle 40 % of teachers; high gains: top 30 % of teachers

References

- Angrist, N., Djankov, S., Goldberg, P.K., Patrinos, H.A., 2021. Measuring human capital using global learning data. *Nature* 592 (7854), 403–408. <https://doi.org/10.1038/s41586-021-03323-7>.
- Araya, M., Rolleston, C., Rose, P., Sabates, R., Tiruneh, D., Woldehanna, T., 2023. Understanding the impact of Large-Scale educational reform on Students' learning outcomes in Ethiopia: the GEQIP-II case, 23/125. *RISE Work. Pap. Ser.* https://doi.org/10.35489/BSG-RISE-WP_2023/125.
- Araya, M., Rose, P., Sabates, R., Tiruneh, D.T., Woldehanna, T., 2022. Learning Losses during the COVID-19 Pandemic in Ethiopia: Comparing Student Achievement in Early Primary Grades before School Closures, and After They Reopened. *RISE Insight Series.* https://doi.org/10.35489/BSG-RISE-WP_2022/049, 2022/049.
- Armor, D.J., Conry-Oseguera, P., Cox, M., King, N.J., McDonnell, L.M., Pascal, A.H., Zellman, G.L. (1976). Analysis of the School Preferred Reading Program in Selected Los Angeles Minority Schools. Retrieved October 20, 2018, from (<https://www.rand.org/pubs/reports/R2007.html>).
- Barnes, A.E., Zuilkowski, S.S., Mekonnen, D., Ramos-Mattoussi, F., 2018. Improving teacher training in Ethiopia: shifting the content and approach of pre-service teacher education. *Teach. Teach. Educ.* 70, 1–11.
- Bau, N., Das, J., 2020. Teacher value added in a Low-Income country. *Am. Econ. J. Econ. Policy* 12 (1), 62–96. <https://doi.org/10.1257/pol.20170243>.
- Bertoni, E., Elacqua, G., Mendez, C., Santos, H., 2024. Teacher selection instruments and teacher Value-Added: evidence from Peru. *Educ. Eval. Policy Anal.* 46 (1), 53–81. <https://doi.org/10.3102/01623737221149417>.
- Brown, C., Andrabi, T., 2023. Inducing Positive Sorting through Performance Pay: Experimental Evidence from Pakistani Schools. *RISE Working Paper Series.* https://doi.org/10.35489/BSGRISE-WP_2023/123.
- Buhl-Wiggers, J., Kerwin, J.T., Smith, J., Thornton, R., 2022. Learning More about Teachers: Estimating Teacher Value-Added and Treatment Effects on Teacher Value-Added in Northern Uganda.
- Cervini, R., 2006. Progreso de aprendizaje en la Educación secundaria Básica de Argentina: un Análisis multinivel de valor agregado. (2018). Reice. *Rev. Iberoam. Sobre Calid. Efic. Y Cambio En. Educ.* 6N. 4 (3). <https://doi.org/10.15366/reice2006.4.3.004>.
- Chetty, R., Friedman, J.N., Rockoff, J.E., 2014. Measuring the impacts of teachers I: evaluating bias in teacher Value-Added estimates. *Am. Econ. Rev.* 104 (9), 2593–2632. <https://doi.org/10.1257/aer.104.9.2593>.
- Cruz-Aguayo, Y., Ibarrarán, P., Schady, N., 2017. Do tests applied to teachers predict their effectiveness? *Econ. Lett.* 159, 108–111.
- Darling-Hammond, L., 2015. Can Value Added Add Value to Teacher Evaluation? *Educ. Res.* 44 (2), 132–137. <https://doi.org/10.3102/0013189X15575346>.
- Dearden, L., Miranda, A., Rabe-Hesketh, S., 2011. Measuring School Value Added with Administrative Data: The Problem of Missing Variables*. *Fiscal Studies* 32 (2), 263–278. <https://doi.org/10.1111/j.1475-5890.2011.00136.x>.
- Dee, T.S., 2005. A teacher like me: Does race, ethnicity, or gender matter? *The American Economic Review* 95 (2), 158–165. <https://doi.org/10.1257/000282805774670446>.
- Eigbiremolun, G.O., Ogbuabor, J.E., Nwamke, C.S., 2020. Estimating private school effects for school children in Peru: evidence from Individual-level panel data. *J. Int. Dev.* 32 (2), 131–149. <https://doi.org/10.1002/jid.3438>.
- Filmer, D., 2023. Long-lived consequences of rapid scale-up? The case of free primary education in six Sub-Saharan African countries. *Int. J. Educ. Dev.* 102, 102872. <https://doi.org/10.1016/j.ijedudev.2023.102872>.
- General Assembly of Pennsylvania (2002). House Bill No. 45, October 9.
- Giffin, J., Hershberg, T., Robertson-Craft, C., 2009. Value-Added as a Classroom Diagnostic in Theodore Hershberg and Claire. In: Robertson-Craft, Eds. (Ed.), *A Grand Bargain for Education*. Harvard Education Press, Cambridge, Mass.
- Gulosino, C., 2018. Evaluating the Tennessee Higher Education Commission's Report Card on the value-added estimates of teacher preparation programs. *Educ. Polic. Anal. Arch.* 26 (0), 33. <https://doi.org/10.14507/epaa.26.2604>.
- Hanushek, E.A., Kain, J.F., Rivkin, S.G., 1998. Teachers, Schools, and Academic Achievement. *NBER Working Paper Series*, p. 6691. <https://doi.org/10.3386/w6691>.
- Hanushek, E.A., Rivkin, S.G., 2012. The distribution of teacher quality and implications for policy. *Annu. Rev. Econ.* 4 (1), 131–157. <https://doi.org/10.1146/annurev-economics-080511111001>.
- Herrmann, M., Walsh, E., Isenberg, E., 2016. Shrinkage of Value-Added Estimates and Characteristics of Pupils with Hard-to-Predict Achievement Levels. *Stat. Public Policy* 3 (1), 110. <https://doi.org/10.1080/2330443X.2016.1182878>.
- Hoddinott, J., Araya, M., Sabates Aysa, R., Woldehanna, T., Tiruneh, D.T., Eryilmaz, N., 2024. Student composition, equity, and mathematics learning outcomes during a time of educational reforms in Ethiopia. *Oxf. Rev. Educ.* 1–19. <https://doi.org/10.1080/03054985.2024.2388047>.
- Hoddinott, J., Iyer, P., Sabates, R., Woldehanna, T., 2019. Evaluating large-scale education reforms in Ethiopia, 19/034. *RISE Work. Pap. Ser.* https://doi.org/10.35489/BSG-RISE-WP_2019/034.
- Isenberg, E., & Hock, H. (2010). Measuring School and Teacher Value Added for IMPACT and TEAM in DC Public Schools. Final Report, 26.
- Iyer, P., Rolleston, C., Rose, P., Woldehanna, T., 2020. A rising tide of access: what consequences for equitable learning in Ethiopia? *Oxf. Rev. Educ.* 46 (5), 601–618. <https://doi.org/10.1080/03054985.2020.1741343>.
- Jerald, C.D. (2009). The Value of Value-Added Data. K-12 Policy. Education Trust. Available at: <https://files.eric.ed.gov/fulltext/ED507719.pdf>.
- Kaffenberger, M., Pritchett, L., 2021. A Structured Model of the Dynamics of Student Learning in Developing Countries, with Applications to Policy. *Int. J. Educ. Dev.* 82, 102371. <https://doi.org/10.1016/j.ijedudev.2021.102371>. ISSN 0738-0593.
- McCaffrey, D.F., & Rand Education. (2003). Evaluating value-added models for teacher accountability / Daniel F. McCaffrey ... [et al.]. RAND Education.
- Meyer, R.H., 1997. Value-added indicators of school performance: A primer. *Econ. Educ. Rev.* 16 (3), 283–301. [https://doi.org/10.1016/S0272-7757\(96\)00081-7](https://doi.org/10.1016/S0272-7757(96)00081-7).
- Munoz Chereau, B., Anwandter, S., & Thomas, S. (2020). Value-added indicators for a fairer Chilean school accountability system: A pending subject.
- Murnane, R.J., 1975. The impact of school resources on the learning of inner city children balinger. Ballinger Pub. Co, Cambridge, Mass.
- National Educational Assessment, Examinations Agency (NEAEA), 2013. Ethiopian 4th National Learning Assessment of grades 4 and 8 pupils. http://213.55.101.25/images/sirna_bar/Grade_4_8_NLA_Reposits_2012.pdf.
- National Educational Assessment, Examinations Agency (NEAEA). Ethiopian 5th National Learning Assessment of grades fourth and eight students. http://213.55.101.25/images/sirna_bar/Grade_4_and_8_NLA_2015.pdf.
- NEAEA, 2022. Early grade Reading assessment (EGRA 2021) report. Ministry of Education, Educational Assessment and Examinations Agency, Addis Ababa.
- Oketch, M., Rolleston, C., Rossiter, J., 2020. Teacher Value-Added using young lives Ethiopia school survey data - diagnosing the learning crisis: potential applications of Value-Added analysis. . 2020/016. *RISE Insight Ser.* https://doi.org/10.35489/BSG-RISE-RI_2020/014.
- Oketch, M., Rolleston, C., Rossiter, J., 2021. Diagnosing the learning crisis: what can value-added analysis contribute? *Int. J. Educ. Dev.* 87, 102507. <https://doi.org/10.1016/j.ijedudev.2021.102507>.

- Perry, T., 2016. English value-added measures: examining the limitations of school performance measurement. *Br. Educ. Res. J.* 42 (6), 1056–1080. <https://doi.org/10.1002/berj.3247>.
- Rolleston, C., Oketch, M., Tiruneh, D.T., Rossiter, J., 2025. Rising access and falling outcomes: educational change and the GEQIP reforms in Ethiopia. *J. Int. Coop. Educ.* <https://doi.org/10.1108/JICE-03-2024-0012>.
- Rossiter, J., Woodhead, M., Rolleston, C., Moore, R., 2018. Delivering on every child's right to basic skills. Young Lives, Oxford. Retrieved from Young Lives website. (<https://www.younglives.org.uk/sites/www.younglives.org.uk/files/YL-Education.pdf>).
- Rowan, B., Correnti, R., Miller, R.J., 2002. What Large-Scale, Survey Research Tells Us About Teacher Effects on Student Achievement: Insights From the Prospects Study of Elementary Schools. *Teach. Coll. Rec.* 104 (8), 1525–1567. <https://doi.org/10.1111/1467-9620.00212>.
- Sabates, R., Singal, N., Teferre, T., Tiruneh, D.T., 2024. Disability and learning in Ethiopia: what changed as a result of the COVID-19 pandemic? *Int. J. Incl. Educ.* 1–20. <https://doi.org/10.1080/13603116.2024.2342362>.
- Scheerens, J., Glas, C., Thomas, S., 2003. Educational evaluation, assessment and monitoring. A systematic approach. Taylor & Francis, Abingdon.
- Shkolnik, J., Hikawa, H., Suttrop, M., Lockwood, J.R., Stecher, B., Bohrnstedt, G., 2002. The Relationship between teacher characteristics and student achievement in reduced-size classes: A study of 6 California districts. In: Bohrnstedt, G., Stecher, B. (Eds.), *What We Have Learned About Class Size Reduction in California: Technical Appendix*. Sacramento, CA: California Department of Education, pp. D1–D22.
- Stacy, B., Guarino, C., Wooldridge, J., 2018. Does the precision and stability of value-added estimates of teacher performance depend on the types of pupils they serve? *Econ. Educ. Rev.* 64, 50–74. <https://doi.org/10.1016/j.econedurev.2018.04.001>.
- Tesfay, N. (2012). Horizontal inequalities in children's educational outcomes in Ethiopia [Master's Thesis. Wolfson College]. <https://doi.org/10.1016/j.ijedudev.2014.08.009>.
- Thomas, S.M., 2020. School and teacher value added performance and the relationship with teacher professional development in mainland China. In: Hall, J., Lindorff, A., Sammons, P. (Eds.), *International Perspectives in Educational Effectiveness Research*. Springer International Publishing, pp. 209–229. https://doi.org/10.1007/978-3-030-44810-3_9.
- Tiruneh, D.T., 2020. COVID-19 School Closures May Further Widen the Inequality Gaps between the Advantaged and the Disadvantaged in Ethiopia. *RISE Blog Ser.* https://riseprogramme.org/blog/COVID-19_ethiopia_school_closures.html.
- Tiruneh, D.T., Hoddinott, J., Rolleston, C., Sabates, R., Woldehanna, T., 2021. Understanding achievement in numeracy among primary school children in Ethiopia: evidence from RISE Ethiopia study. *Res. Improv. Syst. Educ. (RISE)*. https://doi.org/10.35489/BSG-RISE-WP_2021/071.
- Tiruneh, D., Sabates, R., Rolleston, C., Hoddinott, J., 2022. Trends in mathematics learning in Ethiopia: 2012–2019. *Res. Improv. Syst. Educ. (RISE)*. https://doi.org/10.35489/BSG-RISE-RI_2022/045.
- Torres, R., 2018. Tackling inequality? Teacher effects and the socioeconomic gap in educational achievement. Evidence from Chile. *Sch. Eff. Sch. Improv.* 29 (3), 383–417. <https://doi.org/10.1080/09243453.2018.1443143>.
- UNESCO (2024). Ethiopia: Education country brief, available from: <https://www.iicba.unesco.org/en/node/98#:~:text=5Learning%20poverty%2C%20the%20share%20of,unlikely%20to%20achieve%20reading%20proficiency>.
- UNICEF (2024). Ethiopia education cluster quarterly newsletter (2024), available from: <https://reliefweb.int/report/ethiopia/ethiopia-education-cluster-quarterly-newsletter-january-march>.
- Urrea, I. (2018). Teacher value-added in Chile. Tesis para optar al grado de Magister en Economía, Universidad de Chile.
- USAID, 2019. Early grade Reading assessment (EGRA) 2018 endline report. USAID, Addis Ababa.
- Woldehanna, T., Araya, M., Gebremedhin, A., 2016. Assessing Children's Learning Outcomes: A Comparison of two cohorts from Young Lives Ethiopia. *The Ethiopian Journal of Education* 36 (1), 149–187.
- Woldehanna, T., Hagos, A., 2015. Economic shocks and children's dropout from primary school: implications for education policy in Ethiopia. *Afr. Educ. Rev.* 12 (1), 28–47. <https://doi.org/10.1080/18146627.2015.1036548>.
- World Bank (2008). General education quality improvement project (GEQIP-I). Project Appraisal Document. Report No.:45140-ET.
- World Bank (2013). General education quality improvement project II(GEQIP-II). Project Appraisal Document.
- World Bank, 2017. General education quality improvement program for equity (GEQIP-E): project appraisal document: report no: 121294-ET. The World Bank.
- Wright, S.P., Sanders, W.L., Horn, S.P., 1997. Teacher and Classroom Context Effects on Student Achievement: Implications for Teacher Evaluation. *Journal of Personnel Evaluation in Education* 11, 57–67. <https://doi.org/10.1023/A:1007999204543>.
- Yorke, L., Rose, P., Hagos, B., Woldehanna, T., 2020. The effects of COVID-19 on primary education in Ethiopia: Perspectives of school principals and teachers. Research and Policy Paper No. REAL Centre, University of Cambridge.