Oxford Review of Education

ISSN: (Print) (Online) Journal homepage: https://www.tandfonline.com/loi/core20

Expanding educational opportunities or widening learning inequalities? Evidence from national reform of pre-primary education in Ethiopia

Janice Kim & Ricardo Sabates

To cite this article: Janice Kim & Ricardo Sabates (2022): Expanding educational opportunities or widening learning inequalities? Evidence from national reform of pre-primary education in Ethiopia, Oxford Review of Education, DOI: 10.1080/03054985.2022.2072824

To link to this article: https://doi.org/10.1080/03054985.2022.2072824

OXFORD REVIEW OF EDUCATION
https://doi.org/10.1080/03054985.2022.2072824

Routledge
Taylor & Francis Group

Expanding educational opportunities or widening learning inequalities? Evidence from national reform of pre-primary education in Ethiopia

Janice Kim and Ricardo Sabates

Faculty of Education, University of Cambridge, Cambridge, UK

ABSTRACT
Since a nationwide reform of pre-primary education in 2010, Ethiopia has experienced a massive expansion of pre-primary enrolment that increased tenfold in six years. Our paper aims to assess the distribution of early literacy outcomes between children who attended preschool and those who did not and explore how that distribution has changed throughout the reform and by factors such as gender, location, and parental literacy. We find an overall increase in the achievement gaps associated with pre-primary participation between 2010 and 2016. There are also differential patterns in the learning gaps over the reform, with a particular disadvantage for rural students and a relative advantage for students with parents who are not literate. This study suggests that understanding a fuller picture of learning inequality is critical to designing policy to leave no one behind aligned with the UN Sustainable Development Goals.

KEYWORDS
Educational reform; educational inequality; early childhood education; nonparametric statistics; Ethiopia

Introduction

Over recent decades, evidence has shown that investing in early childhood development can yield large economic returns, especially for children from disadvantaged backgrounds (Heckman, 2011; Yoshikawa et al., 2013). In response, global and national commitments towards early childhood education (ECE) have emerged in low- and middle-income countries (LMICs). Small but growing evidence has reaffirmed the unlocked potential of ECE in LMICs, showing that quality early childhood intervention can boost children's developmental outcomes and long-term education, health, and labour market outcomes (Engle et al., 2011; Fernald et al., 2012). The benefits of ECE are consistently greater for the most vulnerable children, thus investment in ECE can be regarded as a cost-effective strategy for reducing inequalities. The UN Sustainable Development Goal's explicit target of providing 'quality early childhood development, care and pre-primary education' for *all* (SDG, Target 4.2, UNESCO, 2015) is a momentous step forward in global policy on ECE.

However, momentum for ECE has not yet developed to ensure equitable access to quality pre-primary education for all. ECE refers to the broad range of organised, site-based educational programmes for children aged three years up to the start of primary

CONTACT Janice Kim jhk50@cam.ac.uk Faculty of Education, University of Cambridge, Cambridge, United Kingdom

education. In our paper, we particularly focus on pre-primary education, an integral component of ECE, which is typically offered for one or two years prior to school entry aiming to develop a variety of the skills children need for school readiness. Strikingly, only one in five children has access to pre-primary education in low-income countries (UNICEF, 2019). In this regard, Ethiopia provides an interesting case to investigate a country's policy response to expanding pre-primary education, which had emerged even before the SDG target was set. In 2010, the government of Ethiopia adopted a new policy framework to promote young children's access to ECE services. The intention of the policy reform was to offer a scalable model of pre-primary education for improving school readiness and educational attainment of children, and to reach those from the most marginalised population (Ministry of Education, 2015). With unprecedented government involvement in pre-primary education, the gross enrolment rates of 4- to 6-year-old children surged from 5% to 50% between 2010 and 2016.

Ethiopia's pre-primary education reform represented a substantial policy departure from a long-standing position to oversee preschool delivered by private or NGO-based providers to large-scale government involvement in public, fee-free pre-primary provision. The main driver of the massive expansion is 'O-Class', a reception classroom attached to primary school for 6-year-olds before they enter Grade 1. Since 2010, O-Class enrolment almost tripled within five years, serving 2.6 million Ethiopian children in 2015/16 – about three-quarters of children who enrolled in preschool. The O-Class is attractive because it can be implemented by expanding a public primary school, which requires limited infrastructure investment and can easily be accommodated within existing structures.

Although there has been an impressive rise in access to pre-primary education, little is known about whether this expansion was done equitably and has addressed learning inequalities between advantaged and disadvantaged students. Evidence from the mass expansion of primary education in LMICs suggests that it has led to more steep wealth gradients related to access, thus being far from accomplishing the goal to reach the most marginalised children (Lewin & Sabates, 2012). Unsurprisingly, the expansion of pre-primary schooling in many low-income countries has followed similar trends, with sharp disparities in access between those who are better off and worse off (Baum et al., 2019; Delprato et al., 2015; Woodhead et al., 2009). A previous study in Ethiopia also describes persistent inequalities in pre-primary access after the 2010 reform by gender, urbanity, and geographical regions (Rossiter et al., 2018).

In addition, a handful of research has documented inequalities in learning outcomes, especially when the education system has undergone rapid transformation. Muralidharan et al. (2019) describe how the entry of a large number of first-generation primary school students into the education system in India as a result of the increase in overall enrolment has exacerbated within-school and within-class inequalities, with students from more diverse backgrounds now going to school. Similarly, Iyer et al. (2020) highlight the case in Ethiopia where access to primary education has more than doubled between 1999 and 2015. With a sudden influx of students from more disadvantaged backgrounds, including the first-generation learners, the massive expansion is likely to be advantageous first for the population with access to resources. As a result, the achievement gaps widen between the first-generation learners and their peers across the primary cycles.

The reform is likely to increase access and probably learning outcomes on average. However, if the rapid reform of pre-primary education is expected to benefit first children from more affluent communities than those from poorer communities, we can hypothesise that there will be widening gaps in both access and learning outcomes between the advantaged and disadvantaged. This paper aims to explore this hypothesis empirically, especially given that equalising access to education will not guarantee more equality in learning outcomes. Rather, as more advantaged groups are likely to be the beneficiaries of the reform, learning inequalities in foundational skills will be widened, in our case between preschool attendees and non-attendees. If the more disadvantaged children remained underserved by existing preschool providers and entered primary school without any learning stimulus, this would be partially responsible for the widening gaps in learning outcomes, particularly for students from disadvantaged backgrounds.

In order to estimate if there have been overall improvements in learning outcomes between students who attended pre-primary school and those who did not, we use the distributional approach. This approach quantifies the dispersion of learning outcomes over the entire sample, as opposed to the traditional statistics like the mean or the median. Of the distributional approaches, we adopt the nonparametric framework recommended by Ho and Reardon (2012) and Ho (2009), as it accounts for the sensitivity of gap estimates to methods, modelling strategies, and psychometrics assumptions. Importantly, it offers transformation-invariant gap estimates across different assessments (Quinn, 2015). In this paper, we leverage individual-level early literacy data to characterise the distribution of foundational skills depending on whether they attended pre-primary school and to show how that distribution is shifted by the expansion of the education system and by factors such as gender, location, and parental literacy. This allows us to track what has happened to learning inequality when the education system has undergone substantial changes driven by the large-scale reform.

Background

Pre-primary education reforms in Ethiopia

Historically, ECE in Ethiopia was provided on a smallscale by kindergartens run by private, non-governmental, and faith-based organisations. Although the importance of ECE for well-rounded child development has been recognised by the government (MoE, 1994), ECE had not been integrated into the public education sectors for the previous two decades. This was partly due to the government's decision on targeted resource allocation to primary and secondary education. As a result, the supply of ECE services remained saturated in urban areas like Addis Ababa, where it served less than 5% of 4- to 6-year-old children, and those it served were from relatively wealthy backgrounds.

New momentum for ECE was formalised in 2010 when the government of Ethiopia developed a National Policy Framework for Early Childhood Care and Education. The policy had four pillars for the delivery of services: parental education; health and early stimulation (prenatal to 3 years); preschools/kindergartens (4–6 years); and community-based non-formal school readiness programmes. Increasing access to and equity in pre-primary education and improving its quality were central in this policy. In particular, the government promoted the expansion of 'O-Class', a reception year for 6-year-olds before

they enter Grade 1. O-Class has been promoted as a low-cost model of service delivery supported by a flexible arrangement between communities and governments (Ministry of Education, 2010; Orkin et al., 2012). In the first year of introduction, O-Class provided one million children with immediate access to pre-primary education, nearly three times as many children as were enrolled in kindergarten the year before (Woodhead et al., 2017). With an increase in the number of children enrolled in O-Class throughout the reform, the gross enrolment ratio (GER) for all 4- to 6-year-olds increased from 5% to 46% over the six years from 2010/11 to 2016/17. The rapid expansion of pre-primary education, however, comes with multiple challenges such as inequitable access to O-Class that favors wealthier communities, insufficient public funding allocated to pre-primary education, and low-quality service provisions which are unlikely to be developmentally appropriate for young children (for more information, see Rossiter et al., 2018).

The current study

Our paper aims to explore whether the large-scale reform of pre-primary education has reinforced or reduced learning inequality in Ethiopia, while it achieves more equality in access to preschool. Using a regionally representative sample of students in Ethiopia, we assess the extent to which the gaps in early literacy outcomes associated with preschool participation have changed over the reform period when the enrolment surged ten times. In addition, we explore differential patterns of the learning gaps across gender, location, and parental literacy. We aim to address the following research questions:

(1) To what extent has the test score distribution of second- and third-grade students' early literacy skills between preschool attendees and non-attendees changed between 2010 and 2016 during the large-scale reform of pre-primary education?
(2) To what extent do estimates of early literacy achievement gaps differ by gender, urban-rural location, and paternal and maternal literacy between 2010 and 2016 during the large-scale reform of pre-primary education?

In this study, we define 'inequality' as opposed to 'equality', which indicates 'the state of being equal in terms of quantity, rank, status, value or degree' (Jacob & Holsinger, 2008, p. 4). Inequality in learning outcomes thus refers to the extent to which children with common characteristics perform differently in a given assessment to measure early literacy. Our second research question focuses more on 'horizontal' inequalities (Stewart, 2002) – inequalities between culturally defined or constructed groups, such as gender, region, language, and ethnicity. Parents' literacy or years of schooling can also be a proxy for horizontal inequalities as it plays a significant role in perpetuating poverty traps across different groups.

A strength of the current study is the ability to leverage two regionally representative datasets on early literacy outcomes to assess how learning achievement is distributed and how these distributions change during the large-scale reform. The study also focuses on children's transition in the early grades (from a reception year through Grades 2 and 3), the focal years in which policymakers and practitioners pay special attention to equipping

children with fundamental skills. Learning in the early years is particularly instrumental in preventing later dropout and grade repetition, thus improving the overall efficiency and effectiveness of the education system (Crouch & Merseth, 2017).

Method

Data, sample, and key variables

The data for this study are drawn from the Early Grade Reading Assessment (EGRA), an influential tool used globally to assess students' early academic ability in reading acquisition (Gove & Wetterberg, 2011). EGRA Ethiopia is a school-based assessment, which was introduced and locally adapted in 2010. We use two EGRA datasets administered in 2010 (pre-reform) and 2016 (post-reform) to 9,121 and 8,332 students in Grades 2 and 3 of primary school, respectively. The EGRA was administered to a regionally representative sample from five regions in Ethiopia, including Tigray, Amhara, Oromia, Somali and Southern Nations, Nationalities, and People's Regions (SNNP, Sidamu language only), which cover 94% of Ethiopia's 4- to 6-year-old population. To ensure regional representativeness, EGRA employed three-stage stratified sampling, using proportional to population sampling at the regional and school levels and stratified sampling at the classroom level by gender and grades. An advantage of using EGRA data is that the surveys straddle the period of pre-primary education reform and were administered to representative samples from the same five largest regions over the two time periods. This enables us to compare learning distributions and its gaps when the system has undergone substantial changes.

In addition, to establish comparability between EGRA 2010 and 2016 assessments, a common-person design (Masters, 1985) was adopted to equate test items from the 2010 and 2016 EGRA instruments, meaning that the same students took part in more than one version of the assessment during the test development process to detect any differences that could be attributed to instrument characteristics, rather than to student characteristics. This process verified the comparability of two test instruments, as well as different administration methods between 2010 and 2016, namely, paper-based versus tablet-based approaches (American Institutes for Research, 2016).

Our analysis includes the following variables from the EGRA datasets:

Pre-primary education participation: As the variable of interest, children were asked to report retrospectively whether s/he had ever been enrolled in pre-primary (i.e. kindergarten, O-Class, or Child-to-Child) classes. We categorised students as having attended pre-primary school if they had attended any of these centre-based or classroom-based programmes, regardless of the type of provision. According to the pre-primary enrolment data from the official education statistics in Ethiopia, in 2010, all preschool participants enrolled in kindergartens as they were the sole ECE provider for the previous two decades. In 2016, about 65% of preschool participants enrolled in O-Class, followed by those enrolled in Child-to-Child (18%) and in private kindergarten (17%; Ministry of Education, 2017).

Oral reading fluency: We used oral reading fluency (ORF) as a measure of early grade reading skills. This approach has been used across different contexts around the world, given that ORF has a high correlation with reading comprehension (Piper & Zuilkowski, 2015). Oral reading fluency is administered to children one-to-one, with an assessor asking a student to read a passage out loud for a period of time, typically one minute. Importantly, ORF has been used to determine ordinal 'reading proficiency' categories in the Ethiopian context (USAID, 2015), which enables us to apply the nonparametric framework for comparing student achievement across different tests.

Child and household characteristics: A range of child and household characteristics, collected from a student's questionnaire, are included as independent variables in the analysis. Drawing on evidence from Ethiopia, on the factors that have been found to be related to pre-primary participation (Vandemoortele, 2018; Woldehanna, 2016), the following characteristics were included: age, gender, father's literacy, mother's literacy, reading materials at home, rural location, and whether the same language is used in the home and at school (see Table 1 for descriptive statistics). In addition, the grade that the child had reached, and the region in which they live, are also taken into account in the model.

Non-response rates were very low in both EGRA datasets across regions and language groups. In EGRA 2010 and 2016, item non-response was almost none across the variables used in the study, from 0% to 0.05%. Little's test on the missing-data mechanism, using Stata command *mcartest*, confirmed that no patterns exist in the missing data in the EGRA datasets. Hence, we employed listwise deletion (i.e. complete case analysis), which is less likely to introduce bias if the data are 'missing at random' and provide accurate estimates of standard errors (Allison, 2002).

Analytical strategy: distribution-wide approach (probability-probability plot)

Our research adopts a distribution-wide, nonparametric approach to summarise achievement gap and gap trends, as proposed by Livingston (2006), Ho (2009), and Ho and Reardon (2012). Using *ordinal* categories of outcomes (e.g. reading proficiency), this approach focuses particularly on the comparison of test score distribution across different tests, which counters the limitations of mean-based gap metrics or percentage-above-cut measures. The mean-based gap metrics rely on the assumption of equal-interval scale properties, which can lead to bias in the estimates if this assumption is violated. The percentage-above-cut has more constraints in applying to cross-test comparisons due to their dependency on cut-score locations (Holland, 2002; Ho, 2009).

Given our research addresses the changes in learning inequalities over time, the approach using a probability-probability plot (PP plot hereafter) allows us to present a transformation invariant comparison of the test-score distributions before and after the large expansion of pre-primary education. The PP plot could be visualised and interpreted as a Receiver Operating Characteristic (ROC) curve of psychology and medicine, which plots the scale score percentiles of one group relative to that of other. The PP plot is thus generated by obtaining all paired cumulative proportions across the score scale underlying the cumulative distribution functions (CDFs). Due to this construction using only paired cumulative proportions and no scale information, all statistics from a PP plot are transformation-invariant. The PP plots, which were first proposed by Wilk and

Table 1. Descriptive statistics of key variables.

	2010		2016		2010			2016		
	Average		Average		Pre (a)	No-Pre (b)	(a)-(b)	Pre (a)	No-Pre (b)	(a)-(b)
	m	(SD)	*m*	(SD)	*m*	*m*	*Diff.*	*m*	*m*	*Diff.*
Preschool Attendance	0.14	(0.35)	0.38	(0.49)	-	-	-	-	-	-
Age	10.14	(2.06)	9.84	(1.68)	9.98	10.17	−0.19*	9.43	10.09	−0.66***
Female	0.50	(0.50)	0.49	(0.50)	0.48	0.50	−0.02	0.51	0.48	0.03**
Father's literacy	0.52	(0.50)	0.72	(0.45)	0.64	0.50	0.14***	0.81	0.66	0.15***
Mother's literacy	0.38	(0.49)	0.47	(0.50)	0.50	0.36	0.14***	0.60	0.38	0.22***
Book at home	0.21	(0.41)	0.44	(0.50)	0.30	0.20	0.10***	0.53	0.38	0.15***
Rural	0.82	(0.38)	0.79	(0.41)	0.66	0.85	−0.19**	0.71	0.84	−0.13***
Same language of instruction	0.90	(0.30)	0.94	(0.23)	0.85	0.91	−0.06	0.94	0.95	−0.01
Oral reading fluency (correct words per minute)	21.78	(21.35)	21.25	(20.93)	23.79	21.42	2.37	23.85	19.66	4.19***
% of proficient reader	0.39	(0.49)	0.39	(0.49)	0.44	0.38	0.06	0.46	0.35	0.12***
Observations	9,121		8,332		1,245	7,876		2,989	5,343	

Note: Sampling weights are used to ensure regional representativeness of the sample. Proficient reader is a combined measure of two benchmarks drawn from four reading proficiency levels – 'reading with increasing fluency and comprehension (Level 3)' and reading fluently with full comprehension (Level 4)' which assesses student exhibiting relatively *functional* reading proficiency levels.

*** $p < 0.01$, ** $p < 0.05$, *$p < 0.1$

Gnanadesikan (1968), were applied in the context of educational test gaps by Spencer (1983), Livingston (2006), Ho (2009), and Ho and Reardon (2012). This nonparametric representation holds both theoretical and practical advantages for cross-test comparisons, particularly as they may help to capture learning outcomes and gaps from a *distribution-wide* perspective (Ho, 2009).

We apply this approach to capturing *'Change in Gaps'* in an ordinal gap measure. First, it compares the achievement gap between preschool attendees and non-attendees in the post-reform period (2016) to the gap in the pre-reform period (2010) using the formula:

$$\Delta G_{ORF} = G_{ORF2} - G_{ORF1} = \frac{P_{2016} - N_{2016}}{SD_{2016pooled}} - \frac{P_{2010} - N_{2010}}{SD_{2010pooled}} \tag{1}$$

where ΔG_{ORF} represents the change in gaps measured by oral reading fluency, P represents the weighted mean of preschool attendees, N represents the weighted mean of non-attendees, and the standard deviations in the denominators are weighted and pooled for the two groups. When both of its component terms are positive, a positive ΔG_{ORF} indicates that the cross-sectional learning gap is wider in 2016 than in 2010, whereas a negative ΔG_{ORF} indicates that the cross-sectional gap is narrower in the post-reform period. When we use an ordinal gap measure, which is expressed as the proficiency categories of oral reading fluency, it is estimated with the V statistics as:

$$V = \sqrt{2}\Phi^{-1}\left(P\left(t_p > t_n\right)\right) \tag{2}$$

where Φ^{-1} is the inverse of the standard normal cumulative distribution function and $P\left(t_p > t_n\right)$ is the probability that a randomly chosen preschool attendee (p) will belong to a higher proficiency level than a randomly chosen non-attendee (n). Because V treats test scores as ordinal, it is invariant to monotonic scale transformations and does not require the interval scale assumption; it requires only the assumption that the test score distributions of the two groups can be transformed to normal (Ho, 2009). When the test score distributions of the two groups are normal, V equals Cohen's d (Ho & Reardon, 2012). When V represents a cross-sectional gap, a gap trend between 2010 and 2016 can be represented as:

$$\Delta V = V_2 - V_1 = \sqrt{2}\Phi^{-1}\left(P\left(t_p > t_n\right)\right)_{2016} - \sqrt{2}\Phi^{-1}\left(P\left(t_p > t_n\right)\right)_{2010} \tag{3}$$

This approach allows us to capture *gap trends* during the policy change with more comparable and comprehensive measures than the mean-based or percentage-based metrics. The goal of this paper is not to estimate the causal effect of some treatment, but rather to characterise the changes in the full underlying distribution of early literacy skills induced by the policy decisions.

The area under the PP curve (AUC), which is equal to PP-based effect sizes, provides a useful framework to measure aggregate test score change and achievement gaps (Bamber, 1975; Ho, 2009). We first render a PP plot by using the Stata command *rocfit*, then use command *roccomp* that provides a chi-squared test for the equality of AUC (DeLong et al., 1988). To account for child and family characteristics, we also use command *rocreg* that can generate a covariate-adjusted PP plot (Janes & Pepe, 2009).

In addition, to check the robustness of our findings using nonparametric graphs and statistics, we estimate the Gini coefficients and the Lorenz curve (Lorenz, 1905), one of the well-known measures to quantify income or wealth inequality from economics. Of the distribution-wide approach, there have been calls to assess learning inequality using the Gini coefficient (Wagner et al., 2018), yet no single measure can be expected to reliably capture all features of a distribution (Rodriguez-Segura et al., 2021). While the PP plot and Lorenz curve share many common properties (Irwin & Hautus, 2015; Schechtman & Schechtman, 2019), the Lorenz curve, which uses the parametric outcomes (i.e. as a continuous measure of oral reading fluency), contains a high dependency on the scale's mean and standard deviation that are often determined by the sample or are arbitrarily selected by the test developers (Lee, 2018). Our aim here is to analyse the trend in inequality through diverse lenses, each with their own strengths and weaknesses, which is likely to provide a fuller understanding of the distributional changes in learning outcomes between preschool attendees and non-attendees over the reform period.

Limitations

Though the EGRA provided a unique opportunity to assess changes in the association between preschool and child outcomes over time, the scope of the analyses and findings are limited by the characteristics of the data. First, although the measure of preschool and all other variables were constructed in exactly the same way across the 2010 and 2016 datasets, a few differences could remain between the two EGRA administrations, such as how to deal with external barriers (e.g. flood, drought, or ethnic clashes) during the sampling and data collection procedure.

Second, EGRA provides a broadly defined measure of children's preschool access, yet it lacks information about pre-primary education programmes such as preschool type, length or duration, class size, teacher–child interaction, language used, and age-appropriate curriculum, among others. It is plausible that, although a region or a sub-group achieved higher pre-primary enrolment during the reform, we are unable to assess the quality of this provision. Given that low-quality pre-primary education is rather detrimental to future developmental outcomes for children, we need additional caution in interpreting the findings in this study.

Third, our analysis is based on measures of foundational literacy and ignores other important outcomes of the learning experience, such as psychosocial and emotional outcomes, socialisation skills, peer interactions, as well as local knowledge. It may be possible that the massive expansion of pre-primary education is associated with many of these wider outcomes which remain unexplored in this paper.

Findings

Sample characteristics and pre-primary education participation: a shift from an elite to a mass system

Table 1 presents descriptive statistics for the sample. On average, children were ten years old at the time of the assessment. About 80% of children were living in rural areas, and among students who attended preschool, the percentage living in rural areas increased

from 66% in 2010 to 71% in 2016. In total, students who attended preschool were more likely to have reading materials at home and live with a father or mother who could read and write than those who did not attend preschool, while overall gaps between the two groups widened from 2010 to 2016. For instance, the differences in mother's literacy between preschool attendees and non-attendees increased from 14 percentage points (50% vs. 36%) to 22 percentage points (60% vs. 38%) between 2010 and 2016. The descriptive picture suggests that children who did not attend preschool have become relatively disadvantaged compared to children who attended preschool. As we saw a substantial change in the composition of preschool attendees and non-attendees between 2010 and 2016 (see, Table 1), we applied a covariate-adjusted ROC curve (Janes & Pepe, 2009) to account for baseline differences of the sample.

With respect to preschool access, approximately 14.2% and 37.9% of children were reported to have attended any form of pre-primary education, nearly tripling between 2010 and 2016. Figure 1 provides further details on the expansion, presenting the distribution of the school-level average of pupils enrolled in pre-primary. This figure depicts how access to pre-primary education in Ethiopia shifted from the *elite* to the *mass* system during the reform. In 2010, more than half of schools (125 out of 237) had less than 10% of students who had ever attended preschool, which shows a skewed right distribution in preschool enrolment. Conversely, in 2016, after the massive expansion of O-Class, the distribution of preschool enrolment shifted close to a normal distribution. About 30% of schools (62 out of 225) reported that more than half of students entered primary school after having attended preschool.

A shift in test score distribution during the reform

Figure 2 presents the gap trends with transformation-invariant scales rendered through the PP plots. The upper panel in Figure 2 of each cohort, F_{pre} (x) and F_{no-pre} (x) denote the proportions of students at or below a given score x on the horizontal axis in the preschool and non-preschool groups, respectively. The test score distributions in the cumulative

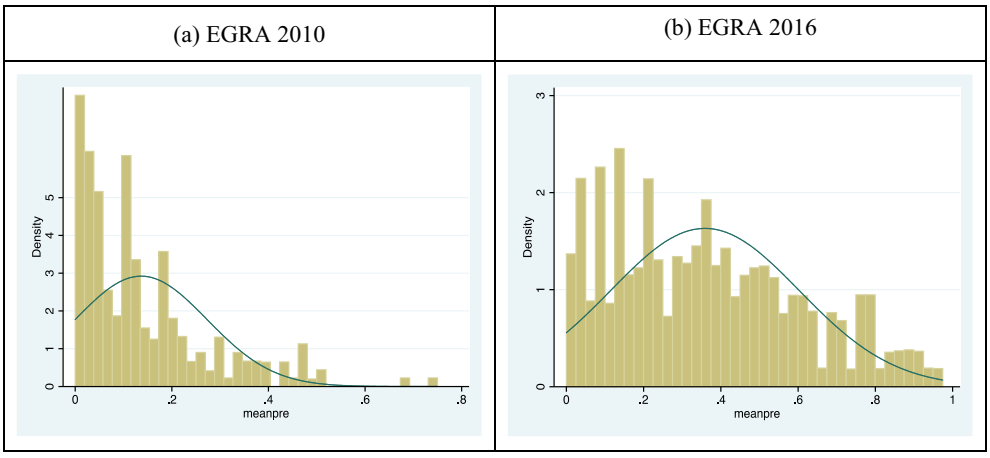

Figure 1. Kernel density of school-level average of pupils who enrolled in preschool.

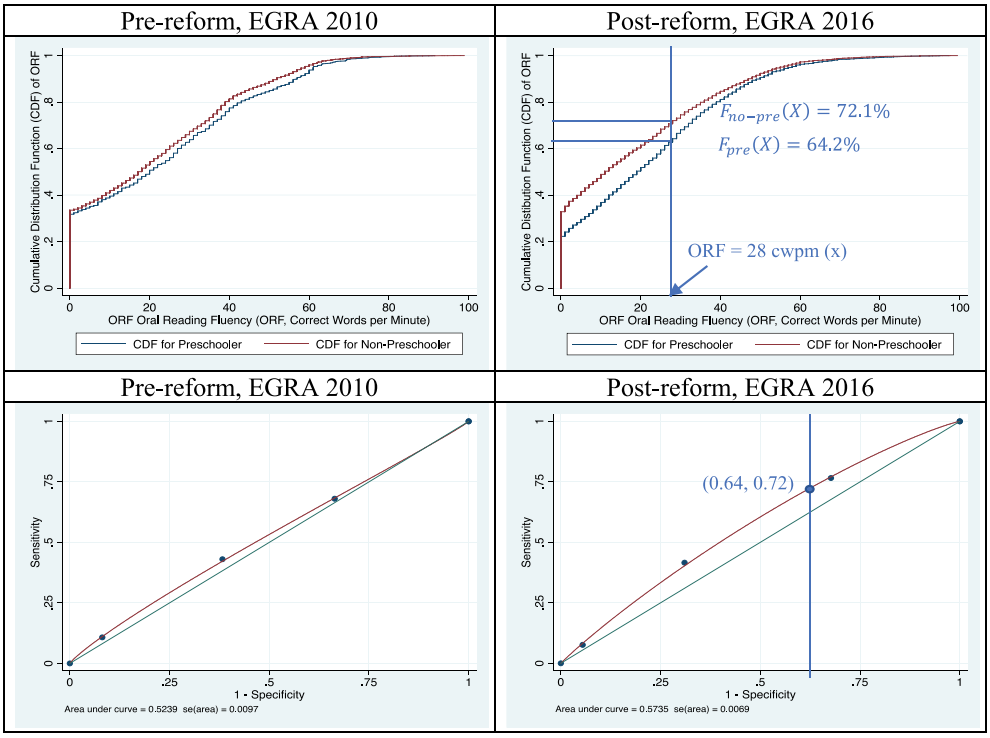

Figure 2. Probability-probability plot: preschool attendees versus non-attendees. Note: The PP plot consists of the cumulative distribution function (CDF, upper panel) and inverse CDF (lower panel) figures. The *rocfit* plot (lower panel) is drawn from the receiver operating characteristic (ROC) curve that displays the full picture of a trade-off between the sensitivity (true positive rate) and (1-specificity) (false-positive rate) across a series of cut-off points. According to Ho (2009), 'each point on the PP plot can be understood as a point plotted from the two intersections of a vertical slice through two CDFs (p. 213)'.

density function (CDF) are generally labelled as a higher-scoring reference distribution, F_{pre} (blue line - visible in online version), and a lower-scoring focal distribution, F_{no-pre} (red line - visible in online version). In the 2016 EGRA cohort, for example, for the average cut score of reading proficiency Level 3 at 28 cwpm (correct words per minute), 64.2% of the preschool group and 72.1% of the non-preschool group was at or below Level 3.

The lower panel is the corresponding PP plot that shows the proportion of the *non-preschool* group below given percentiles of the *preschool* group. This PP plot is generated by obtaining all paired cumulative proportions across the score scale underlying the CDFs (upper panel). The greater the difference between these groups, the further the data point departs from the diagonal line, as expressed by the larger bulge (Livingston, 2006). As stated previously, the area under the curve (AUC) provides a useful measure of achievement gaps.

Compared to the 2010 cohort, results show that all data points for the 2016 cohort lie further to the left of the diagonal line, which denotes the more significant gap between preschool attendees and non-attendees. When we look closely at the AUC, we see that it increased from 0.52 (SE = 0.01) to 0.57 (SE = 0.01) between 2010 and 2016 for the entire

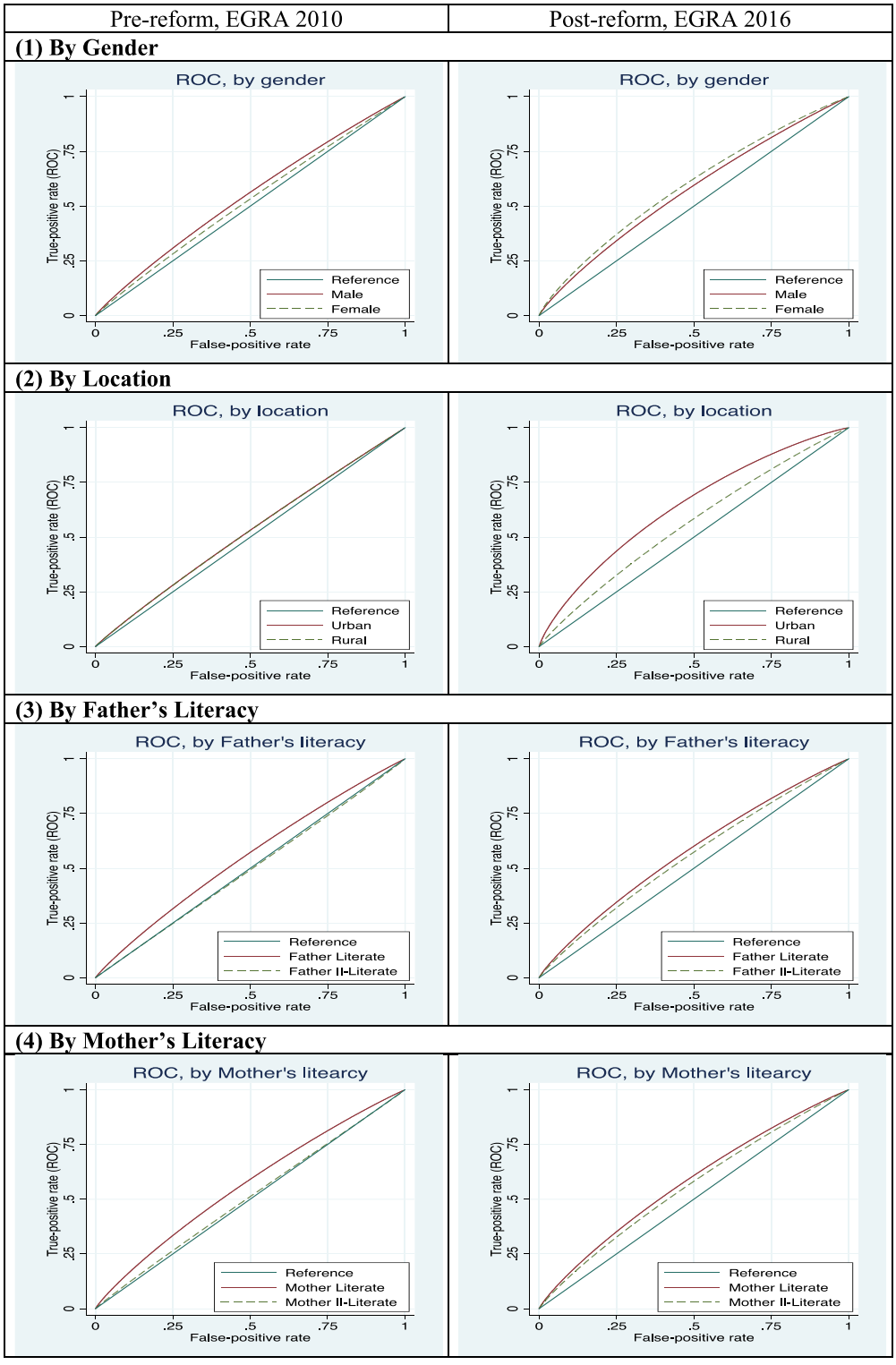

Figure 3. Probability-probability plot by sub-group. Note: The PP plot shows the proportion of the non-preschool group below given percentiles of the preschool group by gender, location, father's literacy, and mother's literacy.

sample. The difference in AUC across the two tests is statistically significant χ^2 (1) = 16.34, $p < 0.001$, suggesting that learning gaps between the two groups widened over the reform periods.

To address our second research question, Figure 3 displays the achievement gaps by sub-groups as defined by (1) gender; (2) urban-rural residence; (3) father's literacy; and (4) mother's literacy. Across all four sub-groups, students who attended preschool outperformed those who didnot attend by a larger margin in 2016 (post-reform) than in 2010 (pre-reform). In terms of the differences by gender, the achievement gaps associated with pre-primary attendance were larger among boys than girls in 2010, however, this gender gap was fully reversed in 2016 as the test score gaps were larger among girls than among boys. Between 2010 and 2016, the AUC increased from 0.51 ($SE = 0.01$) to 0.58 ($SE = 0.01$) for girls, and from 0.53 ($SE = 0.01$) to 0.56 ($SE = 0.01$) for boys. The AUC differences between boys and girls were almost none both in 2010 (χ^2 (1) = 1.84, $p > 0.1$) and in 2016 (χ^2 (1) = 1.86, $p > 0.1$).

When it comes to the gap trends by location, the gaps in test score distribution after the reform are notably larger among children living in urban areas, as opposed to those among children living in rural areas. Stratified by urban and rural residence, the AUC increased from 0.54 ($SE = 0.02$) to 0.64 ($SE = 0.02$) for the urban sample, and from 0.50 ($SE = 0.01$) to 0.55 ($SE = 0.01$) for the rural sample from 2010 to 2016. The AUC difference between the urban and rural sample was 0.04 in 2010 (χ^2 (1) = 3.42, $p < 0.1$), but it doubled to 0.09 in 2016 and became statistically significant (χ^2 (1) = 21.7, $p < 0.001$).

In terms of the gap trends by parental literacy, the learning gaps depending on preschool attendance are larger among children living with literate fathers than those with illiterate fathers in both 2010 and 2016. However, the learning gaps also widen among children living with illiterate fathers in 2016, contrary to these gaps in 2010. Hence, the learning gaps induced by preschool attendance are likely to be reduced between children with literate fathers and those with illiterate parents after the reform. Between 2010 and 2016, the AUC increased from 0.55 ($SE = 0.01$) to 0.56 ($SE = 0.01$) for children living with literate fathers, and from 0.49 ($SE = 0.02$) to 0.54 ($SE = 0.01$) for children living with illiterate fathers. Similar patterns are observed in mother's literacy. Notably, the AUC difference in 2010 was statistically significant by father's literacy (χ^2 (1) = 9.15, $p < 0.001$) and by mother's literacy (χ^2 (1) = 6.66, $p < 0.001$), whereas the AUC difference in 2016 was no longer statistically significant by father's literacy (χ^2 (1) = 1.21, $p > 0.1$) or by mother's literacy (χ^2 (1) = 0.7, $p > 0.1$).

Robustness check

The results using the PP plot hold when we estimate the traditional Gini coefficient using test scores in oral reading fluency. Overall, the Gini coefficients increased from 0.319 to 0.384 between 2010 and 2016. In 2010, the Gini coefficients were 0.316 for preschool attendees and 0.319 for non-attendees, then it increased following the reform to 0.365 for preschool attendees and 0.396 for non-attendees in 2016. Between-group Gini coefficients were also increased from 0.012 to 0.048 over time as a function of preschool attendance. Consistent with the increase in AUC derived from the PP plot, the Gini coefficients tend to increase during the reform, indicating that learning inequalities between the two groups widened between 2010 and 2016.

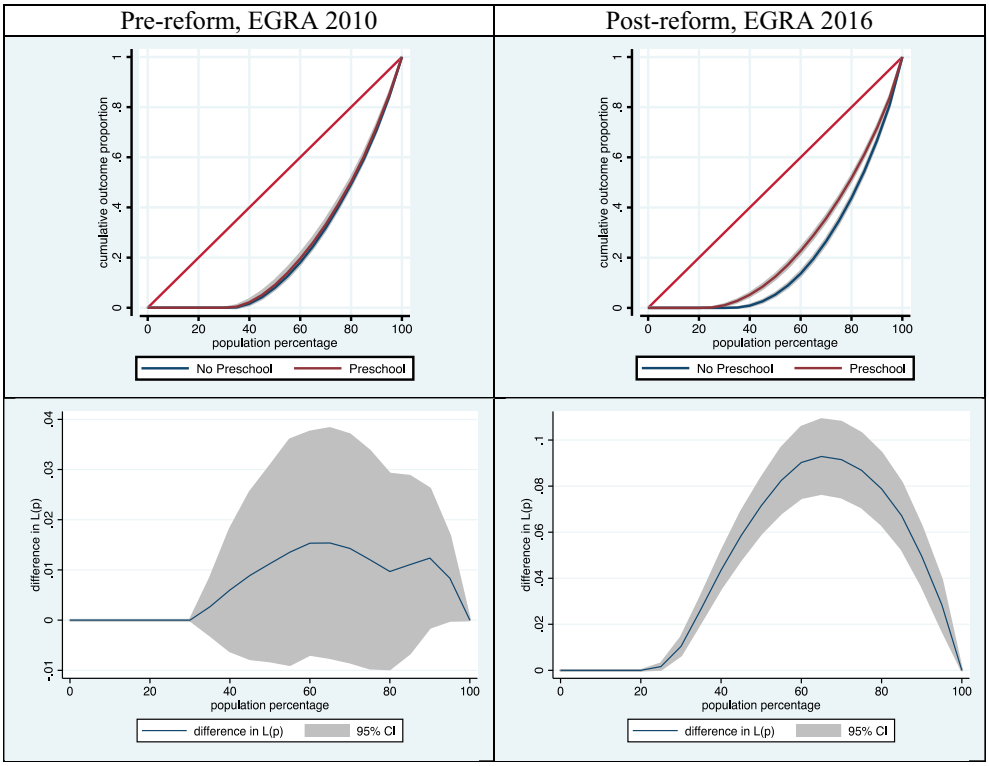

Figure 4. Lorenz curves and contrast plots: preschool attendees versus non-attendees. Note: The contrast plot computes the difference between two Lorenz curves by the percentile of preschool attendees, indicating that the test score distribution for preschool non-attendees is less equal than the test score distribution for preschool attendees.

In Figure 4, the top row panels display the Lorenz curves that compare the distribution of early literacy skills between preschool attendees and non-attendees by each EGRA cohort. It clearly demonstrates that, between 2010 and 2016, students without pre-primary schooling are left further behind than those who attended pre-primary. Thus, a wider learning gap between preschool attendees and non-attendees is partly attributed to the lowest achievement of children from disadvantaged groups who are still

Table 2. Gini coefficient by sub-groups.

		2010			2016		
Gini-coefficient		Preschool	No-Preschool	Between-Group	Preschool	No-Preschool	Between-Group
Gender	Female	0.332	0.324	0.007	0.362	0.391	0.056
	Male	0.301	0.314	0.015	0.364	0.400	0.036
Location	Rural	0.337	0.327	0.003	0.372	0.399	0.031
	Urban	0.278	0.275	0.016	0.321	0.369	0.095
Father's literacy	Literate	0.309	0.320	0.019	0.356	0.383	0.039
	Not literate	0.330	0.317	0.001	0.399	0.424	0.034
Mother's literacy	Literate	0.310	0.335	0.030	0.354	0.372	0.048
	Not literate	0.323	0.310	0.002	0.378	0.410	0.029

underserved by pre-primary providers (i.e. O-Class) despite the large-scale reform. The bottom row panels present the contrast plots from the Lorenz curves that compute the difference between two Lorenz curves by the percentile of preschool attendees. These plots depict a shift in which the underlying distribution moved from a wider dispersion to a narrower dispersion, as learning inequality between preschool attendees and non-attendees increased between 2010 and 2016.

As shown in Table 2, the gap trends in the Gini coefficients by sub-groups show consistent trends with the above results using nonparametric statistics.

Discussion

This study aimed to explore whether the large-scale expansion of pre-primary education has reinforced or reduced the learning inequality in Ethiopia. We assessed the extent to which the distributions of early literacy outcomes between preschool attendees and non-attendees have changed during the reform and explore differential patterns of learning gaps across gender, location, and parental literacy. Our analysis demonstrates that, over the significant policy shift, there was a wider gap of learning distribution in early literacy between preschool attendees and non-attendees. By employing the nonparametric measure of inequality (i.e. PP plot), which offers reliable score metrics for cross-test comparisons, we counter limitations of the *mean*- or *median*-based approach. This reaffirmed that the achievement gaps between preschool attendees and non-attendees significantly widened during the massive expansion of pre-primary education and it may lead to the worsening of learning inequalities.

Why was there a widening learning inequality over the reform? Although beyond the scope of this analysis, we extrapolated a few possible reasons for this finding. First, the current study documents gaps in early literacy skills associated with pre-primary attendance, thus it may relate to *greater selection bias*, especially for those who do not have access to preschool even after the reform. As shown in the descriptive figure (Table 1), the rapid expansion may leave highly disadvantaged children excluded from pre-primary opportunities. Nearly all 4- to 6-year-old children did not attend preschool in 2010; however, half of children from more marginalised communities still do not have access to pre-primary education. Second, in addition to the selection bias caused by family background, preschool attendance was partially explained by supply factors – that is, better-resourced schools may have introduced O-Class earlier than schools with few resources. Due to the cumulative advantages for preschool attendees and cumulative disadvantages for non-attendees induced by the reform, the learning inequalities would be worsened.

Notably, we addressed critical literature gaps by characterising the changes in the full underlying distribution of foundational skills throughout the reform. Understanding how inequality manifests in the broader population can inform whether average-improving policies should be targeted at specific groups, schools and regions, or whether broader reform is needed to have a wider reach (Rodriguez-Segura et al., 2021). To the best of our knowledge, there are only few studies using the distribution-wide approach in LMICs (e.g. Crouch & Rolleston, 2017), while studies mostly come from the US context (see, Livingston, 2006; Quinn, 2015).

We also attempt to substantiate the achievement gap trends by sub-groups, which present mixed results. As for gender, there were no significant learning gaps by gender that have changed during the reform. Only a few studies in LMICs have addressed the gender gap by using the mean-based approaches, and most of them found no significant gender differences in the benefits of attending preschool on academic performance (e.g. Berlinski et al., 2009; Bietenbeck et al., 2017).

We find the urban-rural gap trends associated with preschool increased over the reform period, signalling a significant relative disadvantage for children living in rural areas. Yet, the opposite trends have been observed by parental literacy. The gap trends show that the apparent achievement gaps in 2010 between children with literate fathers/mothers and those with illiterate fathers/mothers were reduced significantly in 2016 despite a sudden influx of children from diverse backgrounds into the education system. This indicates that the learning gaps associated with pre-primary participation have widened between rural and urban children along with the massive expansion, whereas these gaps have narrowed between children with literate parents and those with not literate parents.

Although the urban-rural learning inequalities seem to be growing since the reform, it is encouraging that the gap has been narrowed significantly according to parental literacy. The latter findings could align with the compensatory hypothesis that assumes the benefits of preschool will be greater for the disadvantaged than the advantaged, which has been supported by empirical evidence from LMICs (e.g. Engle et al., 2011). In fact, the findings on the differential effect of pre-primary education by sub-groups remain inconclusive, from none (Bietenbeck et al., 2017) to advantages for children from disadvantaged backgrounds (Berlinski et al., 2009) or more affluent backgrounds (Berg et al., 2013). The gap trends observed in the current analysis point to the importance of intersectionality with gender, urban-rural location, and parental literacy, which all affect the outcomes of the reform for young children. All our analyses are exploratory in nature and do not convey causality. Nevertheless, the results of this study highlight an important point for policymakers that, by understanding how learning is initially distributed and how it evolves throughout the reform, effective policies need to address the resource gaps that end up compounding achievement gaps for different groups of students, especially those from underprivileged communities.

Will the national scale-up of pre-primary education continue to widen the achievement gaps between the advantaged and the disadvantaged? It depends. The new phase of the national-level reform starts shifting its focus from efficiency and effectiveness towards equitable access to high-quality education (World Bank, 2017). Research is needed to assess the effects of these changes on students' learning distribution to critically assess the effects of the government's renewed commitment towards equitable access and learning outcomes for all.

Acknowledgments

We thank USAID Ethiopia for data access and the World Bank (WB) Ethiopia Country Office for the support for the conceptualization of this article. We also thank participants for comments received during the WB's World Development Report 2018 Workshop in Ethiopia, and those attending the annual conference of Comparative and International Education Society 2019 and the UK forum for International Education and Training (UKFIET) 2019.

Disclosure statement

No potential conflict of interest was reported by the authors.

Notes on contributors

Janice Kim :is a Research Associate at the Research for Equitable Learning (REAL) Centre in the Faculty of Education, University of Cambridge. Her research focuses on education reform and policy interventions for school effectiveness and equity, early childhood development, education measurement and evaluation, and education governance and finance.

Ricardo Sabates is a Professor of International Education at the Research for Equitable Learning (REAL) Centre in the Faculty of Education, University of Cambridge. He is a development economist, mainly interested in understanding and measuring educational inequalities and in the application of quantitative methods in education, and contributing to current debates on access and learning for the most marginalised populations.

ORCID

Janice Kim http://orcid.org/0000-0002-4336-0167
Ricardo Sabates http://orcid.org/0000-0002-1433-5667

Data availability statement

Data available within the article or its supplementary materials.

References

Allison, P. (2002). *Missing data.Thousand*. Sage Publications.

American Institutes for Research. (2016). *Early Grade Reading Assessment (EGRA) 2016 midterm report: USAID Reading for Ethiopia's Achievement Developed Monitoring and Evaluation (READ M&E). Commissioned by Addis Ababa: USAID/Ethiopia*.

Bamber, D. (1975). The area above the ordinal dominance graph and the area below the receiver operating characteristic graph. *Journal of Mathematical Psychology*, *12*(4), 387–415. https://doi.org/10.1016/0022-2496(75)90001-2

Baum, D. R., Hernandez, J. E., & Orchard, A. (2019). Early childhood education for all: A mixed-methods study of the global policy agenda in Tanzania. *Early Years*, *39*(3), 260–275. https://doi.org/10.1080/09575146.2019.1572075

Berg, S., Girdwood, E., Shepherd, D., van Wyk, C., Kruger, J., Viljoen, J., . . . Ntaa, P. (2013). *The impact of the introduction of grade R on learning outcomes*. University of Stellenbosch.

Berlinski, S., Galiani, S., & Gertler, P. (2009). The effect of pre-primary education on primary school performance. *Journal of Public Economics*, *93*(1–2), 219–234. https://doi.org/10.1016/j.jpubeco.2008.09.002

Bietenbeck, J., Ericsson, S., & Wamalwa, F. M. (2017). *Preschool attendance, school progression, and cognitive skills in East Africa*. IZA discussion papers, 11212, IZA Institute, Bonn.

Crouch, L., & Rolleston, C., 2017. *Raising the floor on learning levels: Equitable improvement starts with the tail. An insight note from the RISE programme*. Oxford: Research on Improving Systems of Education (RISE) Programme.Retrieved August 2017, from <http://www.riseprogramme.org/sites/www.riseprogramme.org/files/RISE%20Equity%20Insight_0.pdf >

Crouch, L., & Merseth, K. A. (2017). Stumbling at the first step: Efficiency implications of poor performance in the foundational first five years. *Prospects*, *47*(3), 175–196. https://doi.org/10.1007/s11125-017-9401-1

DeLong, E. R., DeLong, D. M., & Clarke-Pearson, D. L. (1988). Comparing the areas under two or more correlated receiver operating characteristic curves: A nonparametric approach. *Biometrics*, *44*(3), 837–845. https://doi.org/10.2307/2531595

Delprato, M., Dunne, M., & Zeitlyn, B. (2015). Preschool attendance: A multilevel analysis of individual and community factors in 21 low and middle-income countries. *International Journal of Quantitative Research in Education*, *3*(1/2), 1. https://doi.org/10.1504/ijqre.2016.073633

Engle, P. L., Fernald, L. C. H., Alderman, H., Behrman, J., O'Gara, C., Yousafzai, A., De Mello, M. C., Hidrobo, M., Ulkuer, N., Ertem, I., & Iltus, S. (2011). Strategies for reducing inequalities and improving developmental outcomes for young children in low-income and middle-income countries. *The Lancet*, *378*(9799), 1339–1353. https://doi.org/10.1016/S0140-6736(11)60889-1

Fernald, L. C. H., Kariger, P., Hidrobo, M., & Gertler, P. J. (2012). *Socioeconomic gradients in child development in very young children: Evidence from India, Indonesia, Peru, and Senegal*. *Proceedings of the National Academy of Sciences of the United States of America (PNAS), 109 (2), 17273-17280*. www.pnas.org/cgi/doi/10.1073/pnas.1121241109

Gove, A. K., & Wetterberg, A. (Eds.).(2011). The Early Grade Reading Assessment: Applications and Interventions to Improve Basic Literacy. North Carolina: RTI International. https://www.rti.org/rti-press-publication/early-grade-reading-assessment

Heckman, J. J. (2011). The economics of inequality: The value of early childhood education. American Educator, 35(1), 31. https://files.eric.ed.gov/fulltext/EJ920516.pdf

Ho, A. D. (2009). A nonparametric framework for comparing trends and gaps across tests. *Journal of Educational and Behavioral Statistics*, *34*(2), 201–228. https://doi.org/10.3102/1076998609332755

Ho, A. D., & Reardon, S. F. (2012). Estimating achievement gaps from test scores reported in ordinal "Proficiency" categories. *Journal of Educational and Behavioral Statistics*, *37*(4), 489–517 doi:10.3102/1076998611411918.

Holland, P. W. (2002). Two measures of change in the gaps between the CDFs of test-score distributions. *Journal of Educational and Behavioral Statistics*, *27*(1), 3–17. https://doi.org/10.3102/10769986027001003

Irwin, R. J., & Hautus, M. J. (2015). Lognormal Lorenz and normal receiver operating characteristic curves as mirror images. *Royal Society Open Science*, *2*(2), 140280. https://doi.org/10.1098/rsos.140280

Iyer, P., Rolleston, C., Rose, P., & Woldehanna, T. (2020). A rising tide of access: What consequences for equitable learning in Ethiopia? *Oxford Review of Education*, *46*(5), 601–618. https://doi.org/10.1080/03054985.2020.1741343

Jacob, W. J., & Holsinger, D. B. (2008). *Inequality in education: Comparative and international perspectives*. CERC Working Paper Studies in Comparative Education, 24. Springer, Dordrecht.

Janes, H., & Pepe, M. S. (2009). Adjusting for covariate effects on classification accuracy using the covariate-adjusted receiver operating characteristic curve. *Biometrika*, *96*(2), 371–382. https://doi.org/10.1093/biomet/asp002

Lee, S. S. (). The gini coefficient's magic does not work on standardized test scores University of Pennsylvania GSE Graduate Student Research 6 . April , 2018 https://repository.upenn.edu/gse_grad_pubs/6.

Lewin, K. M., & Sabates, R. (2012). Who gets what? Is improved access to basic education pro-poor in Sub-Saharan Africa? *International Journal of Educational Development*, *32*(4), 517–528. https://doi.org/10.1016/j.ijedudev.2012.02.013

Livingston, S. A. (2006). Double P-P plots for comparing differences between two groups. *Journal of Educational and Behavioral Statistics*, *31*(4), 431–435. https://doi.org/10.3102/10769986031004431

Lorenz, M. O. (1905). Methods of measuring the concentration of wealth. *Publications of the American Statistical Association*, *9*(70), 209–219. https://doi.org/10.2307/2276207

Masters, G. N. (1985). Common-Person equating with the Rasch model. *Applied Psychological Measurement*, *9*(1), 73–82. https://doi.org/10.1177/014662168500900107

Ministry of Education. (1994). *Education and Training Policy*. St. George Printing Press, Government of Ethiopia. https://planipolis.iiep.unesco.org/en/1994/education-and-training-policy-4042

Ministry of Education. (2010). *National policy framework for early childhood care and education*. Government of Ethiopia.

Ministry of Education. (2015). *Education Sector Development Programme V (ESDP V) 2015/16 –2019/20 G.C*. Government of Ethiopia.

Ministry of Education. (2017). Education Statistics Annual Abstract 2008 E.C. (2017). Addis Ababa: Government of Ethiopia.

Muralidharan, K., Singh, A., & Ganimian, A. J. (2019). Disrupting education? Experimental evidence on technology-aided instruction in India. *American Economic Review*, *109*(4), 1426–1460 https://doi.org/10.1257/aer.20171112

Orkin, K., Yadete, A., & Woodhead, M. (2012). *Delivering quality early learning in low-resource settings progress and challenges in Ethiopia*. The Bernard van Leer Foundation.

Piper, B. (2010). *Ethiopia early grade reading assessment. Data analytics report: Language and early learning Ethiopia*. Addis Ababa: USAID/Ethiopia.https://www.usaid.gov/sites/default/files/documents/1860/Ethiopia_Early_Grade_Reading_Assessment.pdf

Piper, B., & Zuilkowski, S. S. (2015). Assessing reading fluency in Kenya: Oral or silent assessment?. *International Review of Education*, *61*(2), 153–171 doi:10.1007/s11159-015-9470-4.

Quinn, D. M. (2015). Black–white summer learning gaps: Interpreting the variability of estimates across representations. *Educational Evaluation and Policy Analysis*, *37*(1), 50–69. https://doi.org/10.3102/0162373714534522

Rodriguez-Segura, D., Campton, C., Crouch, L., & Slade, T. S. (2021). Looking beyond changes in averages in evaluating foundational learning: Some inequality measures. *International Journal of Educational Development*, *84*, 102411. https://doi.org/10.1016/j.ijedudev.2021.102411

Rossiter, J., Hagos, B., Rose, P., Teferra, T., & Woldehanna, T. (2018). Early Learning in Ethiopia: Equitable access and Learning. *System Diagnostic Report for World Bank Early Learning Program*. https://doi.org/10.5281/zenodo.3371317

Schechtman, E., & Schechtman, G. (2019). The relationship between Gini terminology and the ROC curve. *Metron*, *77*(3), 171–178. https://doi.org/10.1007/s40300-019-00160-7

Spencer, B. D. (1983). On interpreting test scores as social indicators: Statistical considerations. *Journal of Educational Measurement*, *20*(4), 317–333. https://doi.org/10.1111/j.1745-3984.1983.tb00210.x

Stewart, F., 2002. *Horizontal inequalities: A neglected dimension of development*. QEH Working Paper Series – QRHWPS81. Queen Elizabeth House, Oxford, UK.

UNESCO. (2015). *Thematic indicators to monitor the education 2030 agenda*.

UNICEF. (2019). *A world ready to learn: Prioritizing quality early childhood education*.

USAID. (2015). *Results of the early grade reading benchmarking workshop in Ethiopia*.

Vandemoortele, M. (2018). Inequality in attainment from early childhood to adolescence longitudinal evidence from Ethiopia. Oxford: Young Lives. https://www.younglives.org.uk/sites/default/files/migrated/YL-WP177-Vandemoortele.pdf

Wagner, D. A., Wolf, S., Boruch, R. F., Schmelkes, S., Montoya, S., Crouch, L., . . . Van Damme, D. (2018). *Learning at the bottom of the pyramid: Science, measurement, and policy in low-income countries*. UNESCO.

Wilk, A. M. B., & Gnanadesikan, R. (1968). Probability plotting methods for the analysis of data. *Biometrika*, *55*(1), 1–17. http://www.jstor.org/stable/2334448

Woldehanna, T. (2016). Inequality, preschool education and cognitive development in Ethiopia. *International Journal of Behavioral Development*, *40*(6), 509–516. https://doi.org/10.1177/0165025415627700

Woodhead, M., Ames, P., Vennam, U., Abebe, W., & Streuli, N. (2009). *Equity and quality? Challenges for early childhood and primary education in Ethiopia, India, and Peru*. The Hague.

Woodhead, M., Rossiter, J., Dawes, A., & Pankhurst, A. (2017). Scaling-up Early Learning in Ethiopia: Exploring the Potential of O-Class. Young Lives Working Paper No. 163. Oxford: Young Lives. https://www.younglives.org.uk/sites/default/files/migrated/YL-WP163-Woodhead%20%282%29.pdf

World Bank. (2017). Program Appraisal Document, General Education Quality Improvement Program for Equity (GEQIP-E). Washington, D.C.: World Bank. https://documents1.worldbank.org/curated/en/128401513911659858/pdf/ETHIOPIA-EDUC-PAD-11302017.pdf

Yoshikawa, H., Weiland, C., Brooks-Gunn, J., Burchinal, M. R., Espinosa, L. M., Gormley, W. T., Ludwig, J., Magnuson, K. A., Phillips, D., & Zaslow, M. J. (2013). Investing in our future: The evidence base on preschool education. *Society for Research in Child Development and Foundation for Child Development*. https://edsource.org/wp-content/uploads/old/mb_2013_10_16_investing_in_children.pdf