Unpacking the Role of Early Learning in Student Learning Outcomes: Evidence from National Reform of Pre-Primary Education in Ethiopia

Janice Heejin Kim

Emmanuel College

September 2019

This dissertation is submitted for the degree of

Doctor of Philosophy

Faculty of Education - University of Cambridge

ABSTRACT

Unpacking the Role of Early Learning in Student Learning Outcomes: Evidence from National Reform of Pre-Primary Education in Ethiopia

Janice Heejin Kim

Ensuring that young children benefit from their early learning experiences is essential to building a productive and equitable society. Bolstered by accumulated evidence on the high value of investing in the early years, the expansion of early childhood education (ECE) has emerged as a prominent policy agenda in low- and middle-income countries. Nevertheless, substantial gaps remain in our understanding of the conditions that support the scale-up of ECE, particularly in low-resource settings. Responding to this gap, the present dissertation critically examines the role ECE plays in shaping children's educational trajectories by focusing on one country that is experiencing the most rapid and massive expansion of preschool access in sub-Saharan Africa: Ethiopia.

This thesis consists of three distinctive but interconnected papers which assess the relationship between increased access to preschool and student outcomes during the early learning reform initiative in Ethiopia (Chapter 2); the variation in this relationship across a variety of child, family, and school characteristics (Chapter 3); and the sustained benefits of preschool during adolescence (Chapter 4). This dissertation, which is positioned at the intersection of education, economics, and developmental psychology, is theoretically informed by human capital theory (Schultz, 1961; Becker, 1962), and its extension to skill formation technology (Cunha & Heckman, 2007) and bioecological systems theory (Bronfenbrenner, 1979). It draws from the literature on early childhood development and school effectiveness studies.

Chapter 2 examines changes in the relationship between preschool attendance and students' early grade reading achievement during the large-scale expansion of public preschool (O-Class) in Ethiopia. It leverages two Early Grade Reading Assessment datasets that straddle the reform period from 2010 to 2016, during which pre-primary enrolment rates soared by nearly ten times. The results, which are based on ordinary least square/logit regression and school fixed effects, suggest that the large-scale preschool expansion strengthened the role of preschool attendance in predicting second- and third-grade students' reading performance. Overall, patterns before the expansion were overturned by the wider access to preschool. The

reform effort brought significant gains in learning that were substantiated by improved reading test scores, a lower chance of becoming non-readers, and a higher chance of becoming proficient readers. Meanwhile, because each region forged its own expansion plan, the results indicate that there was a huge regional imbalance in the benefits of preschool during the reform period.

Building on the relationships established in Chapter 2, Chapter 3 explores the differential influence of preschool by various child, family, and school characteristics. The results show that, after the ECE expansion, girls benefited more from preschool than boys on their early grade reading achievement. However, preschool benefits were not particularly significant for students from disadvantaged backgrounds, those living in rural areas, or those with illiterate fathers, whereas their advantaged peers obtained greater gains from attending preschool. This is contrary to prior evidence showing that ECE has the ability to reduce learning inequalities. I also found that the relationships between preschool and student outcomes are partially mediated by subsequent school environments. The findings have important policy implications, as the large-scale expansion of preschool may come at the expense of equitable gains between advantaged and disadvantaged children, which could amplify existing learning inequality rather than reduce them. They also call for more attention to ensuring a smooth transition between pre-primary and primary education, as it could affect the sustained benefits of preschool.

Using longitudinal data from the Young Lives Study in Ethiopia, Chapter 4 investigates how the relation between preschool attendance and student outcomes evolves from early childhood to adolescence. The results, which are based on the propensity score matching approach, suggest that preschool attendance led to significant improvement in academic achievement in receptive vocabulary and language and increased educational attainment by age 15, the age at which most students are transitioning to secondary school. However, students from wealthy families likely benefited the most from preschool attendance, while mixed patterns were observed by student gender, parents' education level, and child's prior achievement. Moreover, the results highlight the importance of quality of preschool, as well as subsequent school experiences. These quality dimensions have the potential not only to determine the preservation of preschool benefits but to facilitate students' positive academic trajectories from early

childhood through adolescence. Directions for future research and policy implications related to ECE in Ethiopia are discussed.

ACKNOWLEDGEMENTS

I want to express my sincere gratitude to everyone who has provided support, guidance, and encouragement throughout my doctoral journey. First and foremost, I am deeply grateful to my supervisor, Professor Ricardo Sabates, for setting an example of excellence as a researcher, for pushing me to think critically and creatively, and for sharing his special warmth as a mentor. He has been there for me for each and every step, and his guidance has been invaluable and inspirational. I would also like to thank my advisor Professor Anna Vignoles for supporting me in shaping my initial research ideas and for sharing her critical insights and academic rigour. I have consistently looked to her work for inspiration. I also owe deep gratitude to Professor Panayiotis Antoniou, who has given me thoughtful advice, guidance and encouragement since I embarked on this journey. I would like to thank Professor Pauline Rose for enriching my time at Cambridge. Her exceptional leadership and energy have been inspirational. I would also thank Professor Ben Alcott for generously supporting me as a critical assessor from beginning to end.

Furthermore, I am fortunate to have been part of the wonderful team in Ethiopia, which vastly enriched my experience. I would like to express my sincere gratitude to Hiroshi Saeki, Girma Woldetsadik, Anna Olefir, Fitsum Mulugeta, Tihtina Gebre and other colleagues at the World Bank. I also would like to thank Ato Elias Wakjira, Ato Yasabu Berkneh, Ato Solomon Aberra, and our excellent taskforce team members at the Ministry of Education, Ethiopia, Maekelech Giday at UNICEF, Tassew Woldehanna at Addis Ababa University, and an inspiring ECE expert, Jack Rossiter, all of whose dedication, devotion, and commitment to education made my experience in Ethiopia invaluable. I also owe deep gratitude to my special mentor, Nazmul Chaudhury, for sharing his exceptional insights. I want to thank my friends in and outside of Cambridge, Arif Naveed, Jan Lodge, Nard Choi, Stephen Bayley, Aliya Khalid, Raymond Tangonyire, Catalina Castilo, Ali Yedan, Ju-Yeon, Yu-jin, Mina, and Ji-in, whose companionship always reminds me of the importance of sharing genuinely and generously. I am deeply grateful to Kyu for supporting me tremendously during this adventure.

Last but not least, I feel deeply indebted to my beloved parents, Hyung Joon Kim and Young Sil Choi. Since my childhood, their unconditional love, wisdom, and endless support have been the drive behind my efforts. I've learned from my parents a love and appreciation for

knowledge, a sense of social responsibility, and belief in my true potential. I am blessed to have a wonderful sister and brother-in-law, Hyeon Jin Kim and Si Won Lee, and my beloved niece and nephew, Clare and Justin Lee. Without their support, I wouldn't have gone so far and achieved so much. This dissertation is dedicated to my parents and family.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

LIST OF ABBREVIATIONS

ECE	Early Childhood Education
ECCE	Early Childhood Care and Education
ECD	Early Childhood Development
EGRA	Early Grades Reading Assessment
EMIS	Education Management Information System
ESDP	Education Sector Development Program
GEQIP	General Education Quality Improvement Program
GEQIP-E	General Education Quality Improvement Program for Equity
GER	Gross Enrolment Ratio
KG	Kindergarten
LMIC	Low- and middle-income countries
MOE	Ministry of Education
NEAEA	National Education Assessment and Examinations Agency
NGO	Non-Governmental Organisations
QEAP	Quality Enhancement and Assurance Program
REB	Regional Education Bureau
SDG	Sustainable Development Goals
SEM	Structural Equation Modeling
SNNP	Southern Nation, Nationalities and Peoples (Region)
UNESCO	United Nation Education, Scientific and Cultural Organization
UNICEF	United Nations Children's Fund
USAID	United States Agency for International Development

Note: All dates and years in the dissertation used the Gregorian Calendar.

1 CHAPTER 1 – Introduction

1.1 Context and Motivation

The expansion of early childhood education (ECE) is currently the focus of a prominent policy agenda in low- and middle-income countries (LMICs). The burgeoning evidence accumulated over decades of ECE research has resulted in a global consensus on the value of investing in early childhood development (for reviews, see Black et al., 2017; Engle et al., 2011). Empirical evidence from the fields of neuroscience, developmental psychology, and economics have collectively drawn attention to a child's early years as the most critical and sensitive period in which to establish the foundations of brain architecture (Knudsen, Heckman, Cameron, & Shonkoff, 2006; Shonkoff & Phillips, 2000), and to shape subsequent lifelong developmental potential (Cunha, Heckman, Lochner, & Masterov, 2006; Heckman, 2011). The evidence suggests that investment in ECE is a cost-effective strategy for reducing inequalities perpetuated by poverty and that the benefits are greatest for the most vulnerable children (Engle et al., 2011; Heckman & Masterov, 2007; Magnuson & Duncan, 2017). Bolstered by accumulated evidence, nearly 70 countries currently have national ECE policies, and more countries are in the process of developing them (Richter et al., 2017; Vargas-Barón, 2015). The UN Sustainable Development Goals' explicit target of providing 'quality early childhood development and education' for *all* (SDG, Target 4.2, UNESCO, 2015) is a momentous step forward in global policy on the development of young children.

Despite increased pressure to invest in ECE, evidence on the role of preschool education in promoting learning outcomes remains relatively weak in the LMIC context. This is particularly true in three areas: (1) the effectiveness of a large-scale expansion of pre-primary education, (2) variation in the effectiveness of such initiatives, and (3) the persistent effectiveness of ECE as children progress through formal schooling. The first of these gaps is due to a lack of empirical evidence on the effects of scaled-up ECE programmes or nationwide reforms in pre-primary education, especially in low-income countries (Aguilar & Tansini, 2012; Bastos, Bottan, & Cristia, 2017; Berlinski, Galiani, & Manacorda, 2008; Martinez, Naudeau, & Pereira, 2012; van der Berg et al., 2013). Substantial evidence in the literature of success is largely from small, contextually limited evaluations of early interventions that may not be replicated in ECE programmes operating at scale (Engle et al., 2011; Rao et al., 2014).

There are at least two reasons to recommend caution in extrapolating these findings to large-scale ECE programmes. First, targeted interventions are usually more structured and closely monitored than a universal access programme and often are led by dedicated implementers (Bold et al., 2018; Bouguen et al., 2014). Second, model interventions usually target sub-populations whose responsiveness to the programme may be non-representative (Dumas & Lefranc, 2010). This could be a threat to the external validity of the research. Moreover, as most evidence is from high-income countries, the dramatic difference in resources between rich and poor countries manifests in differences in infrastructure and other education quality indicators, such as adult-child ratio, class size, and teacher qualifications (Engle et al., 2011; Woodhead, 2009).

Understanding the effect of large-scale ECE programmes is also an ‘extremely complicated endeavour’ (Crouch & Destefano, 2017; Phillips et al., 2017), as reflected in the scant number of such studies. Researchers have encountered many challenges in evaluating the average effects of large-scale initiatives, due to numerous factors which are highly contextual and depend on the details of the policies implemented and of the institutional environment (Alderman & Vegas, 2011; Berlinski, Galiani, & Gertler, 2009). The major challenges in isolating the impact of large-scale programmes arise from potential omitted variable biases and endogenous enactment of policies (Ruhm & Waldfogel, 2012). For example, benefits of policies to expand public preschool may be upwardly biased if they are enacted by communities that are already relatively rich in resources for children’s health and development. This challenge is often compounded by the difficulty of obtaining data that are appropriate and sufficiently detailed for this type of evaluation. The empirical evidence in this area is thus limited, and substantial gaps remain in our understanding of the conditions that support the scaling-up of quality programmes.

In the second of these gaps, despite a sudden influx of children from diverse backgrounds into the education system, little is known about variations in the effectiveness of large-scale ECE programmes observed across individual child, family, and school characteristics (Bouguen et al., 2014; Brinkman, Hasan, Jung, Kinnell, & Pradhan, 2017). Given the heterogeneity of the communities in which large-scale ECE programmes are serving young children, understanding the specific contextual sources of differential effects is critically important to education policy (McCoy, Morris, Connors, Gomez, & Yoshikawa, 2016). This is particularly applicable in

LMICs looking to scale-up ECE, as it helps them target their approach to addressing children's needs and to (re)allocate resources to optimise equity.

The third gap in ECE research is that, while there is growing evidence in high-income countries of whether preschool benefits to child outcomes persist in later years (see Ruhm & Waldfogel, 2012, for a review), we know relatively little about the long-term contributions of preschool education in LMICs (Bastos et al., 2017; Woldehanna & Araya, 2017). Prior studies often point to a 'fadeout' of the initial academic benefits of preschool (for reviews, see Claessens, Engel, & Curran, 2014; Gibbs, Ludwig, & Miller, 2011; Bailey et al., 2017), which remains a major puzzle in the literature. This calls for studies that assess the consequences of ECE investment over an extended time span, particularly to determine whether preschool benefits can be sustained in low-resource settings from school entry through adolescence, and into adulthood.

1.2 Purpose, Rationale, and Structure of the Study

This dissertation aims to address these gaps in knowledge of the influence of ECE by focusing on one target country: Ethiopia. Over the past decade, Ethiopia has been experiencing the most rapid and extensive policy reform in pre-primary education in Sub-Saharan Africa. Driven by a large-scale expansion of public preschool initiated in 2010, the ECE landscape in Ethiopia has been transformed from an *elite* system reserved for a few hundred thousand children with affluent backgrounds into the *mass* system that now serves nearly four million young children from all backgrounds. Although impressive progress has been made in preschool access, due to the scarcity of rigorous empirical evidence, it is not clear whether this scale-up initiative actually contributes to achieving the intended policy goals—boosting student outcomes and reducing learning inequalities (MoE, 2015). This dissertation responds to the need for evidence on the early learning reform in Ethiopia; it also has the potential to inform early childhood policies in similar contexts.

This dissertation is driven by the following overarching research question: *What role does early childhood education play in shaping children's educational trajectories?* I focus in particular on the relation between preschool attendance and learning outcomes, and its interaction with children's multi-layered environments that include family, school, community, and policy settings. This dissertation consists of three stand-alone but interlinked essays; each addresses three questions arising from the previous literature. In Chapter 2 (Essay 1), the focus is the *patterns* of the relationship between preschool attendance and students' academic achievement during the massive expansion of pre-primary education in Ethiopia. I examine the extent to which the initial role preschool plays in determining learning outcomes has strengthened or weakened over the reform period. By focusing on the initial period of the reform, I leverage two large, representative datasets of the Early Grade Reading Assessment, which was administered before and after the reform to assess this relationship. In Sub-Saharan Africa, almost no prior research exists on the patterns of how preschool attendance is associated with learning outcomes during a nationwide expansion. This study adds the evidence policymakers, practitioners, and researchers need to consider when deciding whether and how to move forward with such initiatives. Chapter 2 is the most comprehensive chapter of the present dissertation, as it contains the theoretical framework and the context of Ethiopia's education system that are applied to Chapters 3 and 4.

Chapter 3 (Essay 2) focuses on variations in the contribution of preschool for children from various backgrounds. By the extended use of data presented in Chapter 2, I examine whether patterns of the relationship between preschool and students' learning outcomes differ across child and family characteristics, including gender, rural/urban location, parental literacy, and home reading resources. In addition, motivated by prior studies showing that preschool benefits often persist as a function of children attending schools of sufficient quality and support (Bailey et al., 2017; Currie & Thomas, 2000; Magnuson, Ruhm, & Waldfogel, 2007; Puma et al., 2010), I explore the mediating role of subsequent school experiences in the link between preschool and students' learning outcomes. There is little research that illuminates the factors related to promoting equitable gains from attending preschool, thus the present study aims to highlight this important area by exploring whether ECE reforms in Ethiopia helped to reduce inequalities in educational outcomes.

The focus of Chapter 4 (Essay 3) extends to children's outcomes in adolescence, as associated with preschool participation in early childhood. Using a longitudinal dataset drawn from the Young Lives Study in Ethiopia, I examine whether the preschool influence persists in determining children's educational outcomes at age 15. I also further investigate two dimensions explored in Chapter 3—variations across child and family characteristics, and the mediating role of subsequent school experience. Provided that early and later learning prove to be complementary (Cunha & Heckman, 2007), this essay will extend our understanding of the contribution early learning makes to improving learning outcomes and to reducing educational inequalities at a later age.

The present dissertation contributes to the early childhood education research by adding evidence on the reform process and patterns of students' learning associated with investment in early learning in Ethiopia. Consequently, each of the three essays can help to deepen understanding of the scalability and sustainability of early childhood education in the LMIC context. The empirical insights that emerge from this dissertation can facilitate a policy dialogue on how quality early childhood education can reach the most vulnerable children and realise its full potential.

1.3 Definition of Early Childhood Education

Throughout the dissertation, the terms 'early childhood education (ECE)', 'pre-primary education', 'preschool education', and 'early learning' are used interchangeably. These terms refer to formalised early learning provided by centre-based early intervention programmes that foster the physical, cognitive, social, and emotional development of children before they transition to primary school (UNESCO, 2007). The number of years pre-primary education lasts can vary by country, but it usually lasts for two or three years and includes children ages three to seven. In Ethiopia, children can join preschool between the ages of four and six; compulsory education starts at age seven.

While these terms specifically focus on aspects of early learning and education, early childhood care and education (ECCE) and early childhood development (ECD) are broader terms that refer to a comprehensive range of programmes and services provided by multiple sectors. This includes support for early learning (pre-primary schooling and other forms of formal and informal early childhood programmes), stimulation, health, nutrition, water, sanitation and hygiene, and social protection, and covers children from birth until they enter primary school (Naudeau et al., 2011). In the present paper, these broad terms are used when indicating a specific policy document or policy initiative.

1.4 Selection Bias (Endogeneity) of Preschool Attendance

A major challenge in identifying the effects of preschool attendance on students' later outcomes is that selection into pre-primary education is likely not random. Positive selection, whereby parents whose children attend preschool have characteristics that stimulate their child's development and learning, would yield estimates that are biased upward. In Ethiopia, for instance, the key determinants of preschool attendance were belonging to a household with greater wealth, having a more educated caregiver and living in an urban area (Vandemoortele, 2018; Woldehanna, 2016). In the present thesis, to mitigate differences in the demographic, socioeconomic and geographic characteristics of children who selected into preschool versus those who did not, I compare the results of ordinary least square regression analysis to those that hold constant all determinants of preschool attendance and outcomes that do not vary among peers in same school (school fixed effects), or to those that have been matched on observed characteristics (propensity score matching). Although not fully causal, these approaches provide a more rigorous estimate of the associations between preschool attendance and student learning outcomes than a simple comparison of two groups, which fails to adjust for potential sources of confounding bias. In the meantime, I often use conventional terms that include 'effect' for specific empirical approaches–such as marginal effects in a logit regression model and direct and indirect effects in a structural equation modeling–and estimates from these models could be interpreted as association, not as causal inferences.

2 CHAPTER 2 – Preschool and Students' Early Grade Reading Achievement: Have Patterns Changed between 2010 and 2016 during a Large-Scale Expansion of Public Preschool?

2.1 Introduction

In recent years, there has been a notable increase in investment in early childhood education (ECE) in low- and middle-income countries (LMICs). Governments, multinational organisations, and NGOs across the globe are expanding access to ECE, aiming to promote children's holistic development and school readiness, and to use ECE as a strategy to reduce growing inequality (Black, Walker, Fernald, et al., 2017; Engle et al., 2011). These movements have been bolstered by accumulated evidence from neuroscience, developmental psychology, and economics that sheds light on the critical role of early childhood development in enhancing the long-term skill development and health of individuals, and in creating greater benefit for society (e.g., Britto, Yoshikawa, & Boller, 2011; Cunha & Heckman, 2007; Shonkoff & Phillips, 2000). More recently, ECE has been elevated to a prime focus of the UN Sustainable Development Goals, which aim to ensure that *all* girls and boys have access to quality early childhood development, care, and pre-primary education (UNESCO, 2015).

As ECE in LMICs expands, this work needs to be considered in the context of scaling-up early learning. Ethiopia is in the process of a rapid and extensive policy reform in ECE. In the six years after the National Policy Framework for Early Childhood Care and Education was ratified in 2010, the country's gross enrolment rate in pre-primary education surged from 4.8 percent to 50 percent (MoE, 2010; 2016). The Government of Ethiopia's rationale for investment in ECE is to offer a cost-effective model for promoting student learning and grade progression, and to reach the most marginalised communities as an instrument to improve effectiveness, efficiency, and equity in the education system (MoE, 2015). Notably, the provision of public preschool (known as O-Class, a reception year for 6-year-olds that is attached to primary schools) has contributed significantly to achieving greater coverage. With unprecedented public interest in the early years of childhood, the landscape of ECE in Ethiopia has changed substantially, from the private services provided for a few hundred thousand children with affluent backgrounds to the mass system currently serving almost four million young children nationwide.

Despite the rising prominence of ECE in LMICs, and in Ethiopia in particular, little evidence exists on the effects of a large-scale expansion of pre-primary education, and what evidence there is mostly involves high- and upper-middle-income countries (Bastos et al., 2017; Berlinski et al., 2009, 2008; Phillips et al., 2017). The existing large body of research on the positive effects of ECE has relied on small, local, and relatively well-resourced ECE programmes that may not be representative of the service provision that could be afforded at scale (Dumas & Lefranc, 2010; Engle et al., 2011; McCoy, Zuilkowski, Yoshikawa, & Fink, 2017; Rao et al., 2014). Moreover, understanding the effects of a large-scale ECE programme is an 'extremely complicated endeavour' (Phillips et al., 2017). As in any non-experimental intervention or policy initiative operating at scale, there are multiple challenges in measurement, including that average effects may depend on highly contextual factors, such as the selection of participants, institutional environment, and political climate, which are not easily or readily observable (Alderman & Vegas, 2011; Berlinski et al., 2009; Ruhm & Waldfogel, 2012). Substantial gaps thus remain in our understanding of the conditions that support the scale-up of quality programmes, particularly in low-resource settings.

To fill this gap, the present study focuses on the *patterns* of the relationship between preschool attendance and students' academic achievement during a massive expansion of pre-primary education. This study examines the extent to which the initial role of preschool in predicting learning outcomes for students in Ethiopia was *strengthened* or *weakened* during the period of reform, especially their performance in basic literacy. Given that the Government of Ethiopia has been embarked on a nationwide O-Class expansion in 2011, I leverage two regionally/ linguistically representative datasets of the Early Grade Reading Assessment (EGRA), which was administered in 2010 (pre-reform) and 2016 (post-reform) to assess whether the relationships have changed over time. This corresponds in particular to the six-year period of the reform, during which the participation of Ethiopian children in pre-primary education soared to nearly ten times the pre-reform enrolment rates.

The present study adds to the limited literature on scaling-up early learning in the LMIC context. To the best of my knowledge, this study is the first to rigorously examine the relationship between preschool attendance and students' outcomes during a large-scale expansion of pre-primary education in Sub-Saharan Africa using the latest representative data. My findings not only deepen our understanding of how the role of preschool has evolved through the reform, they also generate policy insights for resource allocation and next steps in providing quality ECE for *all* girls and boys. The rest of this chapter is structured as follows: I summarise the relevant literature in Section 2, introduce a theoretical framework in Section 3, provide the background of Ethiopia's education system and recent early learning reform in Section 4, before I set out the purpose and research questions of the present study in Section 5. I describe the data in Section 6, provide the empirical methods used in this study in Section 7, followed by the results in Section 8. I provide more discussion and note the limitations of this study in Sections 9 and 10, and conclude in Section 11.

2.2 Literature Review

The current research is related to two areas of literature, both of which are relatively limited but growing. The first relevant body of empirical literature addresses the effects of early childhood education in LMICs. Broadly, I provide general findings from multi-country studies, regional studies in Sub-Saharan Africa, and prior studies conducted explicitly in Ethiopia. The second relevant body of literature focuses on the effects of scaling-up early childhood education, which covers studies from both high-income and LMICs. Finally, I discuss the current study's contribution to the literature by adding evidence on the scale-up of early childhood education in LMICs.

2.2.1 Empirical Evidence on the Effects of Early Childhood Education in LMICs

Research on ECE in LMICs has generally identified the positive effects formal ECE has on individual child development. Recent reviews of studies summarising the estimated effects of ECE across highly diverse settings in LMICs include a 2011 article in the *Lancet* that reviewed nine ECE studies (Engle et al., 2011) and a recent meta-analysis of 26 ECE interventions (Rao et al., 2014), both of which presented consistent evidence of the benefits of ECE attendance for children's cognitive outcomes. For the purposes of the present study, this review is focused on the role of ECE in predicting academic achievement and educational attainment by comparing ECE attendees and non-attendees. [1] Academic achievement in this case is

[1] Thus, the reference category for preschool attendees is preschool non-attendees (otherwise, as indicated).

understood as children's learning outcomes, as measured by achievement tests of academic skills such as reading and mathematics.

First, two studies on the influence of preschool in rural Bangladesh found that first-grade students who had attended preschool showed better performance in reading, writing, and oral math than those who had not (Aboud & Hossain, 2011; Aboud, Hossain, & O'Gara, 2008). In rural Cambodia, children who had any preschool experience—home-based, community-based, or state-run preschool—performed significantly better on pre-academic skills and motor skills than those who had not participated in any programmes; those attending state-run preschools made higher gains than those in home- or community-based preschools (Rao, Sun, Pearson, et al., 2012). Relatedly, first-grade students in rural China who attended kindergarten or independent pre-primary classes demonstrated higher performance on literacy and math assessments than those who had no preschool experience (Rao, Sun, Zhou, & Zhang, 2012).

Similar to the present study, Gove et al. (2018) examined the link between preschool participation and early literacy skills in 16 countries, as measured by the EGRA.[2] Data were collected between 2008 and 2016 from national and regionally representative EGRA surveys, with most of the datasets focusing on students in second and third grade. As a result of linear regression, only 5 of the 16 countries showed a positive and statistically significant association between preschool and oral reading fluency, from 1.4 to 4.7 more words per minute in Ghana and Tanzania (both $p < .01$) to 12.8 more words per minute in Indonesia ($p < .001$).[3] A simple correlation plot using the preschool enrolment rate among the study population in each country indicated a positive relationship between overall participation in pre-primary education and preschool influence on early grade reading outcomes (Gove et al., 2018). This study included the EGRA 2014 dataset collected in Ethiopia; however, this was a pilot survey for only two mother tongues in one region (Hadiya and Wolaytta in SNNP) and thus is not comparable with the sample in the present study.

There is limited but growing evidence on ECE from Sub-Saharan Africa. In rural Mozambique, a randomised experimental study of a community-based preschool programme found that, two

---

[2] These include Ghana, Indonesia, Iraq, Jordan, Kenya, Malawi, Mali, Nicaragua, Nigeria, Philippines, Rwanda, Senegal, Tanzania, Uganda, Zambia, and Ethiopia.

[3] Linear regression model includes a set of control variables: gender, urban/rural location, language of instruction, and socioeconomic status (based on the average of child-reported articles in the home into wealth index quartiles).

years after the start of the programme, children were more likely to be enrolled in primary school and had higher cognitive skills in language and early math (Martinez et al., 2012). In a Madrasa preschool programme in Kenya, Zanzibar, and Uganda, children who had attended 18 months of preschool had higher levels of school readiness in verbal, non-verbal, and numeric aspects of cognition than children who had not attended preschool (Mwaura, Sylva, & Malmberg, 2008). The study in rural Uganda documented the positive and lasting association between preschool attendance and academic achievement in math for sixth-grade students (Hungi & Ngware, 2018).

Two fairly recent studies examined the predictive role of ECE on a national sample in Sub-Saharan Africa. A study in Zambia, which used child development data collected as part of a national assessment of 6-year-olds, found that preschool participation resulted in improved school readiness across multiple development domains, such as receptive vocabulary, letter naming, and nonverbal reasoning (McCoy, Zuilkowski, et al., 2017).[4] Using data from the UWEZO survey conducted with a nationally representative sample in Kenya and Tanzania,[5] Bietenbeck, Ericsson, & Wamalwa (2017) reported small but statistically significant benefits of preschool attendance on children's academic achievement in literacy and numeracy across students age 7 to 16, benefits their peers with no early learning experience did not show.

As for empirical evidence from Ethiopia in particular, only a handful of papers have addressed the relationship between preschool participation and children's cognitive development in an urban context. The Young Lives Study found that children who attended preschool between 2006 and 2008 showed better performance in receptive vocabulary and math assessments, which was sustained up to age 8 (Woldehanna, 2011, 2016; Woldehanna & Gebremedhin, 2012); however, the gains from preschool faded out by age 12, as measured by math achievement (Vandemoortele, 2018). In a small-scale evaluation of Ethiopian preschool treated by the Emergent Literacy and Math programme in the Oromia region, Dowd, Borisova, Amente, and Yenew (2016) found that children who attended any preschool of standard or

[4] Martinez et al. (2012) and McCoy et al. (2017) reported on cognitive, socioemotional, and motor skills using a comprehensive child development assessment, such as IDELA.

[5] UWEZO is the nationally representative household survey of school-age children's education, and their literacy and numeracy skills in Kenya, Tanzania, and Uganda.

enhanced quality showed significant improvement in early literacy and numeracy skills, which did not occur among children who did not attend preschool.

The studies presented above collectively indicate that there is strong evidence to support the claim that ECE in LMICs and Sub-Saharan Africa had a positive influence on children's academic skills in primary school; however, they relied mostly on small-scale and contextually limited trials. Although ECE study quality remains variable (Rao et al., 2014), these studies applied rigorous experimental (e.g., Martinez et al., 2012) and quasi-experimental designs (e.g., Bietenbeck et al., 2017; McCoy et al., 2017) to elucidate the role of preschool attendance in improving students' learning outcomes.

2.2.2 Empirical Evidence on the Effects of 'Scaling-Up' Early Childhood Education

Evaluation of the effects of scaled-up ECE programmes on students' learning outcomes is highly relevant for policymakers who are considering large-scale, widely accessible early learning programmes. However, limited empirical evidence on how these policies affect child outcomes remains a concern in both developed and developing countries (Engle et al., 2011; Ruhm & Waldfogel, 2012). Taking into account that the present study focuses on a large-scale expansion of preschool, this section is of particular relevance to understanding the scope of what effects can be expected from scaled-up preschool initiatives on academic achievement and educational attainment in both high-income countries and LMICs.

U.S. and high- or upper-middle-income countries. Recent studies from the U.S. and high- or upper-middle-income countries demonstrate that participation in scaled-up preschool programmes yields sizable short-term benefits in literacy and math achievement in primary school (see Phillips et al., 2017; Ruhm & Waldfogel, 2012; Wong, Cook, Barnett, & Jung, 2008) The evidence that emerged from the U.S. focuses on statewide universal pre-kindergarten (pre-K) programmes that serve about 30 percent of the nation's 4-year-olds (Phillips et al., 2017). In a universal pre-K programme in Georgia, Fitzpatrick, (2010) found that preschool attendance increased fourth-grade reading and math test scores, especially for disadvantaged children residing in rural areas and small towns.[6] An evaluation of Oklahoma's

[6] Since these are intent-to-treat estimates, many of the results are driven by higher enrolment gains in rural areas relative to urban areas (Fitzpatrick, 2010).

universal pre-K found substantial gains in children’s pre-reading and pre-writing skills, such as spelling and word identification (Gormley, Gayer, Phillips, & Dawson, 2005), and significant benefits from preschool attendance were observed among Hispanic students (Gormley, 2008). Studies conducted in New Jersey and North Carolina consistently documented the positive association between universal pre-K programmes and students’ academic achievement in reading (language) and math from Grade 3 to Grade 5 (Barnett, Jung, Youn, & Frede, 2013; Ladd, Muschkin, & Dodge, 2014).

Apart from studies on statewide pre-K policies, Bassok, Gibbs, and Latham (2018) explored nationwide patterns in the U.S. between 1998 and 2010 of how preschool participation was associated with higher literacy and math scores at school entry and in Grade 3. While the magnitude of these associations quickly faded as children progressed through school (e.g., the preschool advantage lasted to Grade 1 for literacy and to Grade 3 for math), patterns in 2010 mirrored those in 1998, despite the fact that this period featured heightened public interest in ECE in the U.S. (Bassok et al., 2018). A few studies that examined the long-term patterns of preschool expansion showed that these benefits can persist into adolescence and adulthood (for the U.S., see Cascio, 2009; Ludwid & Miller, 2007; for France, see Dumas & Lefranc, 2010; for Norway, see Havnes & Mogstad, 2011). For example, Cascio (2009) found that, by exploiting the state-by-state expansion of kindergarten in public schools in the 1960s and 1970s, children affected by the pre-K expansion were less likely to drop out of high school or to be incarcerated later in life, although these benefits were found only for white students.

Several papers have documented the impact of preschool expansion in other countries, highly concentrated in Latin America. Focusing on a massive, government-led preschool construction programme in Argentina, Berlinski et al. (2009) found that increased access to preschool had significant effects on children’s academic achievement on language and math tests three years after participating in a public preschool. Similarly, Berlinski et al. (2008) explored a rapid expansion of public preschool in Uruguay and found significantly positive effects of preschool attendance on school attainment by age 15, which was equivalent to 0.8 additional years of schooling. More recently, evaluation of a large-scale expansion of pre-primary education in rural Guatemala found that preschool attendance moderately improved grade progression at the proper age for 12-year-old students (Bastos et al., 2017).

In the Sub-Saharan African context, South Africa experienced a massive expansion of the Grade R provision, which was similar to the O-Class reform in Ethiopia. Between 2001 and 2012, preschool enrolment rates in South Africa doubled from 39 percent to 78 percent. Using the National Assessment results in 2012, Berg et al. (2013) reported that preschool attendance translated into small positive outcomes; for example, 50 additional days in a 200-day school year for languages and only 12 days for mathematics in Grade 5. In terms of the differential benefits of preschool by school wealth quintiles, Grade R had no measurable benefit for students from the poorest three school quintiles, whereas students from the wealthiest quintile schools experienced statistically significant positive influence. The evidence from South Africa suggests that, without sufficient attention to the quality of teaching and learning during the expansion of pre-primary education, the equity gap may be widened and expanded, to the advantage of children attending more affluent schools (Berg et al., 2013).

Taken together, these studies that applied a credible non-experimental evaluation strategy suggest that scaled-up preschool initiatives have the potential to deliver important short- and long-term benefits in high- and upper-middle-income countries.[7] However, this evidence cannot easily be extrapolated to poorer countries, where far fewer children have access to preschool; where infrastructure, financial, and human resources are limited; and where the quality of pre-primary school may be considerably lower than in developed countries. These factors may combine to yield different results than the average formal ECE programme in a relatively well-resourced setting.

Lower-middle-income countries. Some unique empirical evidence from lower-middle-income countries drawn from a randomised controlled experiment in Cambodia (Bouguen et al., 2014) and a non-experimental evaluation in Indonesia (Brinkman et al., 2017) suggests that the effects of increased access to preschool can be highly context specific. Both studies evaluated

[7] Various non-experimental methodologies were applied to the reviewed studies. (1) Difference-in-difference (DID): Fitzpatrick (2008); Ladd, Muschkin, and Dodge (2014); Cascio (2009); Berlinski, Galiani, and Gertler (2009); Berlinski, Galiani, and Manacorda (2008); Bastos et al. (2017); Havnes and Mogstad (2012); Brinkman et al. (2017); (2) Regression discontinuity design (RDD): Gormley et al. (2005, 2008); Ludwig & Miller, (2007); Barnett, Jung, Youn, & Frede, 2013; (3) Instrumental Variables (IV): Dumas and Lefranc (2012); Brinkman et al. (2017); (4) Sibling fixed effect: Berlinski, Galiani, and Manacorda (2008); and (5) School fixed effect: Berg et al. (2012). Because the focus of the present study is on pre-primary education (school- or centre-based ECE), comprehensive early childhood interventions with cross-sectoral elements, such as nutrition, health, parental education, and pre-primary education (e.g., Jamaica by Getler et al., 2015, and Columbia by Bernal et al., 2012) were not included in the review.

a large-scale expansion of preschool coverage in villages situated in especially disadvantaged rural areas. The scale of the interventions and the fact that both were implemented by the government, instead of by a dedicated NGO, makes this a compelling case for an evaluation with potentially high external validity in LMICs (Bouguen et al., 2014).

First, focusing on an extensive preschool construction programme in rural Cambodia, Bouguen et al. (2014) found a *negative* short-term effect of preschools on the cognitive development of children. The most significant negative effects were found among the children of poorer and less educated parents. These alarming results are partly attributed to severe implementation constraints that led to limited exposure to preschool and poor service quality. Notably, the findings highlight the importance of understanding parents' behavioural responses to new preschool initiatives relative to the alternative (existing) options available. To illustrate, for some children, mostly those from wealthy backgrounds, parental response led to a choice to participate in preschool as a substitute for underage enrolment in primary school. By contrast, for those largely from poorer backgrounds, parental response led to withdrawal from any formal education, since stricter policies on the minimum age for primary school entry was enforced. Cambodia's case exemplifies that positive results can be hard to replicate for large-scale programmes led by the government, which require a careful design that accounts for implementation conditions and behavioural responses to an intervention in a given context (Bouguen et al., 2014).

Work by Brinkman et al. (2017) in Indonesia presents how a government-sponsored, community-based preschool (mostly delivered by playgroups) affected child development in a rural context. While the intervention led to broader and longer preschool participation for all, the benefits of preschool were conditioned on its duration and the household characteristics of the child. Among children exposed to preschool for at least three years, versus those who never attended preschool, the authors found a modest and sustained impact on children's language and cognitive development, especially for those from disadvantaged backgrounds. This finding suggests that preschool is likely to supplement the limited household resources or poor parenting practices of underprivileged children. Similarly to the evidence from Cambodia (Bouguen et al., 2014), the results from Indonesia highlight the importance of understanding the full context of policy implementation and behavioural responses from parents and communities such as substituting existing services (e.g., kindergartens) for new preschool programmes, and communities' imperfect compliance with the intended interventions (Brinkman et al., 2017).

In sum, results from evaluations of scaled-up programmes were more variable (Engle et al., 2011) than the results from small-scale model programmes. This implies that the implementation capacity, constraints, and policy environments of a particular system are significant drivers of success or failure. Empirical evidence also shows that the results from ECE programmes that reached large and representative populations were generally smaller than those found for small-scale programmes (Duncan & Magnuson, 2013). Even though it is not possible to generalise these findings to the Ethiopian ECE programmes, these studies provide a comprehensive perspective on what could cause variations in the effects of preschool within a country and the scope of effects expected from scaled-up initiatives that could be explored by the present study.

2.2.3 Contribution to the Existing Literature

The present study contributes to this literature in two significant ways. The first contribution is that, to the best of my knowledge, it is the first study of students' academic achievement associated with the large-scale expansion of preschool in Sub-Saharan Africa. Given that the composition of preschool attendees likely will be changed considerably by an influx of previously excluded children, it is paramount to understand how patterns of the relation between preschool attendance and child outcomes have changed over the course of the reform. To achieve this, I leveraged two representative datasets that capture the period spanning a major shift in the education system from serving primarily the most advantaged children to including a wider representation of society.

The second contribution is that this study presents evidence of the influence of expanded access to preschool in a context not studied before. Much of the previous evidence on large-scale preschool programmes was from countries where at least half of preschool-aged children already had access to preschool, and this may yield different results than the average ECE programme in poorer countries with limited access. Moreover, unlike the cases in Cambodia and Indonesia which had (donor's) financial support, a rapid expansion of preschool in Ethiopia has been rolled out without a sufficient financial commitment, targeted resourcing, or capacity-building for regional and local ECE experts. Impressively, the construction of O-Classes, that reach about 75 percent of primary schoolers nationwide, is driven by contributions from families, communities, and local governments. This may mirror actual conditions and practical challenges that many Sub-Saharan African and other low-income countries are currently facing while they consider how to move forward with the large-scale, universal programmes proposed by the UN Sustainable Development Goals.

2.3 Theoretical Framework

Theories of human development from both economics and developmental psychology offer useful lenses for investigating the relation between preschool attendance and children's skill development in the context of policy change. The human capital theory in the field of economics highlights that, to the extent that early and later learning are complementary (Carneiro & Heckman, 2003; Cunha & Heckman, 2007), a lack of education in the early years reduces children's efficiency in lifelong learning once they enter school. The bioecological theory from developmental psychology emphasises that child development is a process that unfolds over time through interactions between a child's individual characteristics and the context in which the child belongs (Bronfenbrenner & Morris, 2006; Bronfenbrenner, 1979, 1986). While human capital theory allows us to justify that early childhood development is the most cost-effective form of human capital investment for building a productive and equitable society, the bioecological theory allows us to understand contexts that galvanise public interest in ECE, rapid changes in the system, and systemic adaptation to change, which are embedded in the multi-layered environments.

In the present section, I start with a brief overview of the main tenets of the two theories and introduce a combined framework that serves as a lens for exploring the link between variations in the policy environment and variations in individual child developmental trajectories. This review is followed by a discussion of how these theories complement each other and enhance our understanding of the role preschool attendance plays in determining the academic abilities of Ethiopian children, which were potentially strengthened or weakened during the early learning reform.

2.3.1 Human Capital Theory from Economics

In human capital theory from the economic literature, human development is the result of accumulated investment over one's life course. Stemming from the work of Becker (1967, 1975) and Schultz (1961), human capital theory frames education as an investment that individuals and families make to increase their stock of knowledge and gain skills that ultimately will yield higher future earnings and success.[8] Human capital theory generates various forms of the production function for a child's cognitive skills (e.g., Leibowitz, 1974; Todd & Wolpin, 2003). For example, child cognitive development is considered a knowledge-acquisition process in which current and past inputs are combined with an individual's genetically endowed ability to learn to produce a current cognitive outcome (Todd & Wolpin, 2003).

Extended from the traditional theory of human capital, Carneiro and Heckman (2003) and Cunha and Heckman (2007) theorise a model of skill formation technology. They view skill development as 'an interactive and multi-stage process of promoting multi-dimensional skills', where children who benefit from early human capital investments may benefit more from later investments. Two key insights from skill formation technology are established: that skills are self-productive and complement each other, and that skills complement investment.

First, skill formation technology suggests that human capital accumulation results from the 'self-productivity' of multiple skills—cognitive skills, non-cognitive skills, and health—developed in earlier stages which bolster the development of these skills at subsequent stages. Second, it highlights the 'dynamic complementarities' of skills by suggesting that skills formed in one life stage raise the productivity of investment at subsequent stages over the life cycle (Cunha & Heckman, 2007; Heckman, 2006), and that future skills have intergenerational impact (Heckman & Mosso, 2014). These dynamic relationships, jointly called 'skills beget skills', make early life a critical period, as it is the time when children lay the foundation for building skills later in life (Cunha et al., 2006).[9] Combined with neurobiological perspectives,

[8] Earlier literature points out that the returns to investments made in early childhood are likely to be higher than returns to investments made later in life, simply because beneficiaries have a longer time to reap the rewards (Becker, 1964).

[9] However, some critiques point out that the hypothesis of dynamic complementarity in early childhood rests on a thin empirical base (Magnuson & Duncan, 2016). Evidence using more recent data from an experimental

it is clear that the returns to human capital investments made in early childhood—the most critical and sensitive period of formation of the brain's architecture for learning and development—are much larger than those made in later life (Knudsen et al., 2006).

The notion of dynamic complementarity, taking on greater importance in the early years, has significant implications for the political and financial commitment to ECE in LMICs (Engle et al., 2011; Naudeau et al., 2011). From the equity point of view, investing in disadvantaged young children is a policy with no 'equity-efficiency trade-off', as it convincingly reduces the inequality induced by poverty and raises the productivity of society at large (Cunha & Heckman, 2007; Heckman & Masterov, 2007). The theoretical and empirical evidence conveys a clear message that, if a lack of early stimulation and learning reduces human capital investments or adversely affect an individual's level of human development, the costs of missing this critical period may accumulate over time, particularly for at-risk children. Remediation of a loss of early learning, while not impossible, is comparatively expensive and may impose substantial burdens on society (Knudsen et al., 2006).

The human capital theory framework provides many strengths that lead to empirically testable hypotheses (i.e., education production function) about the effects of early investment on children's outcomes. Economic models in particular emphasise how individuals and families make investment decisions to achieve a set of goals, under constraints imposed by the relevant budgets, physical environments, network, and policies (Wuermli et al., 2012). To illustrate, the expansion of fee-free preschool affects the decision-making process, partly by altering the external conditions and constraints (e.g., distance to preschool) under which households operate, and partly by changing the resources people need to send their children to preschool.[10] The economic model also provides major criteria for policy evaluation, as well as methodological approaches to address several challenges in the empirical estimation (Wuermli et al., 2012). For instance, it highlights the importance of accounting for unobserved variables, selection bias, and various measurement problems to prevent biased estimates. In the present study, the human capital theory offers a structure for mapping the family background and

evaluation of Head Start does not find that significantly larger gains accrue to students who enter the programme with higher skill at programme entry (Purtell & Gershoff, 2013).

[10] The family's preferences and beliefs are also important determinants of early investment (Cunha, 2013).

current skill level of the individual child, along with investment in the child's development and growth at an early stage.

2.3.2 Bioecological Theory from Developmental Psychology

Apart from understanding the potential relationship between investment in early childhood education and learning outcomes, the present study is primarily interested in the early learning reform in Ethiopia. The conceptual premise for the different ways a major shift in education systems can result in benefits for this relationship are complex. To enhance our understanding of the child developmental process through interaction with the multi-layered environments surrounding them, the current study uses *bioecological* theory advanced by Bronfenbrenner (1979, 1986) and Bronfenbrenner and Morris (2006). [11]

The developmental psychology literature provides comprehensive conceptual and operational definitions of human development (see Baltes, Lindenberger, & Staudinger, 1998; Gottlieb, Wahlsten, & Lickliter, 2007; Lerner, 1986, 1998; Sameroff, 1983). This approach views child development as a set of processes that are inextricably linked to the multiple contexts and systems children inhabit. An individual's growth in character and competence is not achieved in isolation but is, rather, shaped by continuity and change in families, schools, peers, neighbourhoods, and broader social contexts. As an earlier application of 'dynamic systems' theory to child development (e.g., Von Bertalanffy, 1968; citing Yoshikawa et al., 2018), Bronfenbrenner's seminal work conceptualises child development that occurs within the bioecological systems through dynamic interactions between individuals, their families, and their environments. This approach views individual lives as surrounded by a set of nested structures and differentiates between the *proximal* environment that is directly experienced by the individual (e.g., the family context) over an extended period and the more *distal* cultural, social, and economic systems that have an indirect effect on the individual, which are often mediated by the more proximal context (Bronfenbrenner, 1979, 1986).[12] The advantage of the

[11] Earlier in his work (1979, 1986) it was called the ecological theory, but Bronfenbrenner continues to revisit and expand his own theory. In his latest work (Bronfenbrenner, 1995; Bronfenbrenner & Morris, 2006), the term 'bio'ecological theory stresses that 'living organisms whose bio-psychological characteristics, both as a species and as individuals, have as much to do with their development as do the environments in which they live their lives' (Bronfenbrenner, 1995, p. 8).

[12] The distinction between proximal and distal factors and their meanings can depend on the context in which they are being applied.

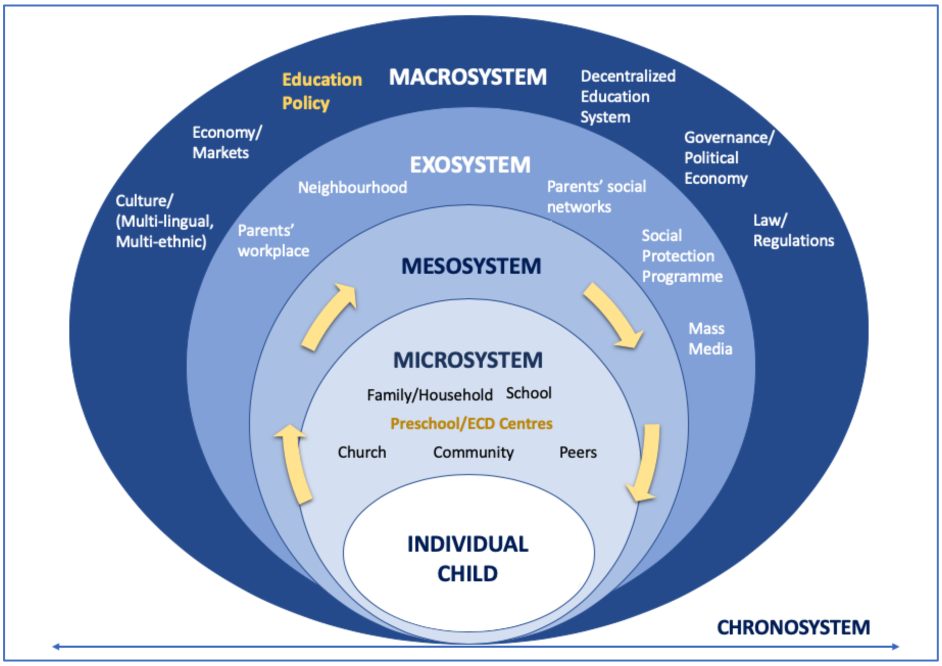
bioecological approach is that it allows us to focus on the interaction between the interrelated instrumental factors rather than on a simple list of influences, and so provides a useful framework for understanding systemically how multi-level factors influence individual development.

According to Bronfenbrenner's bioecological theory, four defining elements affect child development: (1) process, (2) person, (3) context, and (4) time. Forces deriving from multiple settings (e.g., family, peers, schools, and communities), from the relations among those settings, and from the individual's relation to and interactions with those settings all contribute to child development (Bronfenbrenner & Morris, 2006). Centred on the individual child as a primary force of development, his or her interactions with and between the surrounding context —such as family, school and community environment (micro-systems), the larger society, networks and institutions (exo-systems or meso-systems), and culture, economics, and policies (macro-systems)—support, sustain, or hinder growth.[13] Within this multi-layered system, bioecological theory highlights that changing any one of the four elements (process, person, context, time) of interactions with the surrounding systems can change the individual's development trajectory.

Policies in particular determine an important part of the context in which families make decisions about child development, what Bronfenbrenner (1979, 1986) calls the macro-system. A central question on how bioecologies influence child development is how macro-system contexts and events (e.g., national reform in pre-primary education that prioritises universal access to public preschool) influence intermediate contexts (e.g., policy responses from decentralised regional and local governments). It, in turn, influences the settings or contexts within the developing individual's microsystem (families, schools, peers, and neighbourhood), the settings within which the person has face-to-face interactions or proximal processes. To illustrate, the large-scale expansion of preschool, which induces a shift from the elite system to the mass system, is thought to affect the bioecology of human development by changing the macrosystem, as depicted in Figure 2.1.

[13] It should be noted that, in bioecological theory, the direction of influence between children and their surroundings is not a unidirectional but a bidirectional or reciprocal relationship. Thus, while environment influences child development, a child interacts with it and can change the way, for example, how parents treat or respond to their needs.

Figure 2.1. A Bioecological Model of Human Development



Source: Adaptation of Bronfenbrenner's (1979, 1986) bioecological model of child development

Drawing from bioecological theory (Bronfenbrenner & Morris, 2006), I hypothesise that, if there is a major shift in the education system, the relationships between preschool and an individual's development, which are nested within this macrosystem, will be changed in a way that strengthens or weakens the role of preschool in determining a child's skills. In response to policy changes, parents' decision to send their children to preschool will be made in different circumstances that interact with both proximal family factors (e.g., parenting style, educational aspirations, and language use) and distal family factors (e.g., parental education, family size, income and poverty, and maternal employment), all of which are crucial for directing the growth of a child. In the present study, the time component, which usually indicates different periods or transitions in the life course, points to the importance of historical times and cross-sectional events at the macro level.

Yoshikawa and Hsueh (2001) note that the particular benefits of a bioecological (dynamic systems) perspective is that 'the principles are chosen to highlight change over time in the policy environment, child's development, and systems intervening between the two' (p. 1,889). This approach, which stresses the reciprocity of influence among the components of systems, may serve as an interface between policy implementation and the diversity of families. Despite its potential role in bridging public policy and child development (Ford & Lerner, 1992; Thelen

& Smith, 1994), the knowledge gap in 'linking policy variation to developmental variation' has not yet been addressed by the comprehensive frame, especially in large-scale policy implementations (Yoshikawa & Hsueh, 2001; Yoshikawa et al., 2018).

Therefore, the present study attempts to underscore the ways in which extrafamilial contexts might affect intrafamilial processes and individual variation. In particular, my first research question in the present study (Chapter 2) addresses the linkage between policy variation and developmental variation through the novel gap metrics using students' learning distribution. In Chapters 3 and 4, this perspective is applied further to address a set of potential moderators and mediators that may help to explain the pathways through which ECE participation affects child development. Nevertheless, bioecological theory has the limitation that it has a very high demand for data; for example, various forces from the multi-layered environments within which children grow up for empirical testing of the model (Wuermli et al., 2012); thus, the present study used it as a conceptual underpinning that enriches our interpretation of the results in a given context.

2.3.3 Interdisciplinary Framework: Complementing Two Theories

Both human capital theory and bioecological theory provide theoretical links between early learning and child development, pointing out that there are diverse contexts in which early educational experiences are converted into gains in skill formation. However, in the meantime, each theoretical explanation suggests a slightly different focus. First, with respect to the dynamic nature of human development, the human capital model directs attention to *interaction* between investments made in multiple stages of child development (e.g., parental investment between initial conditions (t-1) and intervening period (t)), while the bioecological model directs attention to the *interaction* between the individual and his or her multi-layered environment over time.

Second, these two models lead to contrasting predictions regarding the *complementary* versus *compensatory* role of early investment in child development (Magnuson et al., 2016; Wuermli et al., 2012). In human capital theory, the hypothesis of dynamic complementarity (Cunha & Heckman, 2007) implies that the benefits of preschool may be amplified if children possess high skills upon preschool entry and when followed up by high-quality, enriched environments

(Aizer & Cunha, 2012).[14] From this point of view, early childhood education plays a significant role in preventing equity gaps as early as possible before the disparities widen over the life course (Currie & Thomas, 1999; Heckman & Mosso, 2014). In contrast, developmental theories focus on a compensatory (or substitute) role of preschool in which the benefits of an enriched early learning environment are most pronounced for children with the least prior exposure to such environments (Ramey & Ramey, 1998; Watamura et al., 2011). This view links the productivity of early investments to the match between the qualities of ECE programmes and the specific developmental supports a child needs, especially on the programmes' responsiveness to individual risk, impairment, and adversity (Blair & Raver, 2012). While complementary and compensatory models lead to different education production predictions, both imply that preschool benefits will diverge according to the characteristics of the child, family, preschool, subsequent schooling, and broader policy environments.

Third, in terms of how each theoretical approach has influenced public policy, human capital theories emphasise the role of economic efficiency and distribution as guides to policy. Policy changes can 'significantly alter the context in which individuals make human capital investments, effectively shifting the marginal benefits and marginal costs for private human capital investment decisions, thus changing the optimal level of these investments for individuals and families' (Wuermli et al., 2012, p.44). By comparison, human developmental theories stress interacting environments and contexts in guiding policy decisions 'by identifying complex mechanisms of action through which a programme may affect particular outcomes (i.e., mediators), and by providing insights as to why a programme may operate differently in different settings or for different target sub-groups (i.e., moderators)' (Wolf, Aber, & Morris, 2013, p. 3). Both theoretical approaches have contributed to guiding policy decisions—for example, cost-benefit analysis, optimal resource allocation, and targeting a marginalised group—related to child development.

In the present study, an interdisciplinary framework enabled me to delve into the human developmental processes in a context of systems change. It provided theoretical and empirical support for my argument that a massive expansion of preschool is a pivotal factor in reshaping

[14] The downside to this complementary is that it is often difficult to recover from early deficiencies. Later investments can complement previous ones but likely may not substitute completely for earlier missing components.

the relationships between early learning and child development, viewed as interactions in complex systems. Using this framework yielded a broader understanding and a more comprehensive set of analytical tools to link policy variation to human development variation, as addressed in the current study.

2.4 Background and Context: The Case of Ethiopia

2.4.1 General Context of Ethiopia and the Education System

Ethiopia is a country in the Horn of Africa with a population close to 100 million. As a region, Sub-Saharan Africa has the largest proportion of young children experiencing malnutrition and poverty (Black et al., 2017) and not reaching basic developmental milestones (McCoy, Peet, et al., 2016). On the Human Development Index, a composite statistic of life expectancy, education, and income per capita indicators conducted by the UNDP, Ethiopia ranks near the bottom, number 173 out of 189 countries (UNDP, 2018). According to the prevalence of stunted growth, the indicator used to assess a child's physical and cognitive development potential, 38 percent of Ethiopian children under five are stunted in size, with widespread variation in the degree of stunting between and within regions (CSA, 2016).[15]

Since the 1990s, Ethiopia has made remarkable progress toward achieving the Education For All goal of universal primary education. In 1992, almost four out of five children were out of school. Two decades later, in academic year 2015-2016, net enrolment rates reached 100 percent (MOE, 2016).[16] The rapid expansion of access to primary school has been driven by strong leadership and a commitment from the Government of Ethiopia, in collaboration with international development partners. Abolishing school fees, constructing new schools closer to where children live, training new teachers, promoting parental and community involvement, and embracing previously marginalised groups have all contributed to the success of universal primary education (Engel, 2011). Over the past decade, the proportion of the education budget out of the total government budget has remained a steady 20 percent (Khan et al., 2014).

[15] Stunting is measured by a height-for-age z-score of more than 2 standard deviations below the World Health Organization Child Growth Standards median (WHO, 2009), which show a restriction of a child's potential growth (Black et al., 2008).
[16] The net enrolment rates by gender were 104 percent for boys, 95 percent for girls.

In 2009, the General Education Quality Improvement Programme (GEQIP), a multi-donor-supported programme to improve the learning environment in schools, commenced, which was a key objective under the country's third and fourth Education Sector Development Plan (ESDP). GEQIP I (2009-2013) and GEQIP II (2014-2018) have addressed the essential elements for improving students' outcomes by improving the supply and deployment of qualified teachers; providing textbooks, learning materials, and teacher training; distributing capitation school grants; and establishing an inspection system of school quality (World Bank, 2017). Although learning outcomes have shown some improvement during this period, internal inefficiency, inequity, and poor education quality remain persistent challenges in Ethiopia's education sector. Completion rates in primary education have stagnated at 50 percent for a decade (MoE, 2016); particular groups, including girls, students with special needs, and those from pastoralist communities, are still excluded from access to quality education; and student learning outcomes have remained very low—about 44 percent of Grade 4 students tested nationally were at or below the basic level in reading (World Bank, 2016, 2017). Importantly, the entrenched problem of the dropout rate in Grade 1 reaching 18 percent (MOE, 2016), as well as the lack of basic academic skills, calls for more attention to the 'foundational grades', including pre-primary and primary, which can equip children adequately for future schooling and learning.

2.4.2 Historical Review of Ethiopian Early Childhood Education: Reform Process

Formal compulsory primary education in Ethiopia starts at the age of seven. Hence, pre-primary education targets children four to six years old. Historically, ECE in Ethiopia has been provided on a small scale by private, non-governmental, and faith-based organizations. Ethiopia's 1994 Education and Training Policy document initially acknowledged the provision of ECE for the 'all-round development of the child in preparation for formal schooling' (MoE, 1994). Nevertheless, in the last two decades, ECE has not been integrated into the public education and health sectors. This is partly due to the government's decision on resource allocation to selected areas, which directed significant resources to basic primary and secondary education (MoE, 2002).[17] The government encouraged the private-sector and

[17] In 2002, the Ethiopian Ministry of Education explicitly stated that it did not have the resources to focus on preschool education, preferring rather to consolidate the primary school system: '[…] from the perspective of Ethiopia's economic capacity, the opening of kindergartens involving massive expenditure cannot be a top

longstanding faith-based suppliers to provide ECE services (Hoot, Szente, & Mebratu, 2004). As a result, the supply of ECE services remained saturated in towns and cities where it served less than 5 percent of 4- to 6-year-old children, and those it served were from relatively wealthy backgrounds and lived predominantly in urban areas.

Creating the will: Political change and emerging actors. Since 2007, a national constituency advocating for the expansion of accessible and affordable ECE has emerged. There were several key drivers for this initial movement. First, as a result of global pressure to achieve Education For All, concern about unequal early learning opportunities was raised among government officials in Ethiopia (Rossiter, Hagos, Rose, Teferra, and Woldehanna, 2018). The Education For All goals encouraged achieving equitable service provision by 'expanding and improving comprehensive early childhood care and education, especially for the most vulnerable and disadvantaged children' (UNESCO, 2000). Second, with longstanding support from UNICEF, the political leadership of the state minister for general education was instrumental in seeking a multi-sectoral early childhood care and education (ECCE) strategy. Without clear sectoral guidelines for ECCE, service provision for young children was 'inadequate, but also fragmentary and lacking in coordination' (MoE, 2010b, p. 18). A policy network was created by forming the ECCE Task Force that included government officials from three ministries (Ministry of Education, Ministry of Health, and Ministry of Women's and Children's Affairs), UNICEF, NGOs, civil society organizations, and academic partners (Addis Ababa University and Kotebe University), which contributed to the coordination of sectors to achieve an integrated approach to ECCE.

In 2007, the first diagnostic study of the pre-primary education sector was conducted to identify major challenges and opportunities in the delivery of ECCE. These challenges included high fees; a lack of training and poor working conditions for teachers; lack of an adequate, culturally sensitive curriculum; and a lack of awareness of the importance of early learning among families and communities (Orkin, Yadete, & Woodhead, 2012). The analysis also identified a number of existing opportunities, including private investors, NGOs, and religious

priority, as regular universal primary education has not yet been achieved. Thus, the opening of kindergartens is an area that has been left for private investors and religious organisations, and for parents who can afford to pay the fees' (MoE, 2002).

organizations with an interest in ECCE; and local rural institutions, such as women's and farmers' associations that could support the implementation of ECCE services.

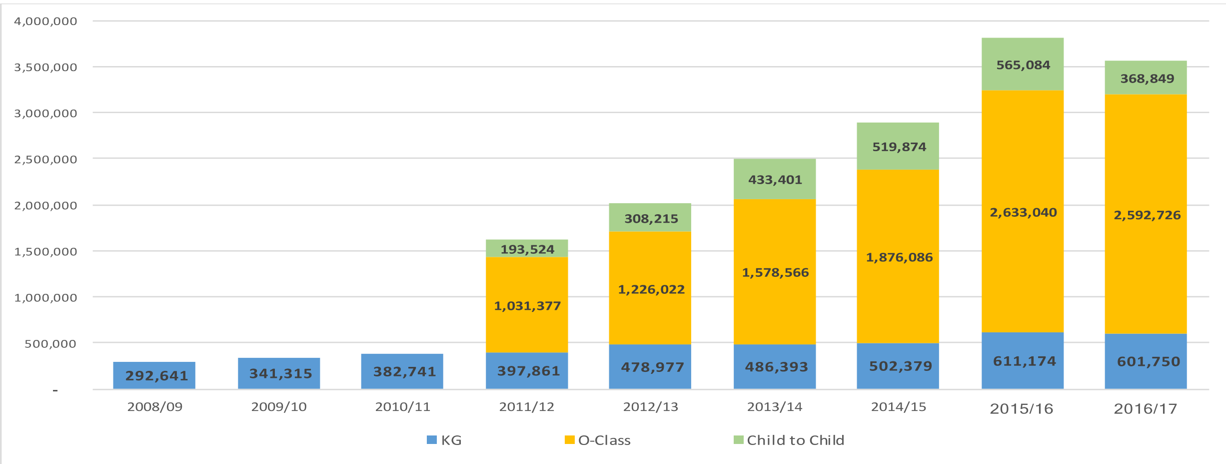
In 2008 and 2009, new early learning initiatives emerged, including a pilot of the Getting Ready for School Child-to-Child (CtoC) approach and the first signs of a reception class, which was acknowledged as 'an interesting initiative [that] has been launched with success: it consists of organizing a pre-primary class within an existing primary school. This has helped spreading ECCE into rural areas' (MoE, 2010a). These initiatives helped to demonstrate the potential of early learning to improve school readiness and revealed a strong demand for early learning services, especially among communities in rural areas (AIR, 2013; Rossiter et al., 2018). This encouraged policymakers and international actors to work closely on issues related to early childhood development.

***Policy reform for early learning: Transition to the new mass system*.** New momentum for ECE was formalised in 2010 when the Government of Ethiopia developed a National Policy Framework for ECCE. The National Policy Framework is to provide a holistic and comprehensive approach to the development of children from the prenatal period to seven years of age (MoE, 2010b). The framework was developed through a participatory and multi-sectoral process spearheaded by the Ministry of Education, in coordination with the Ministry of Health and Ministry of Women's Affairs (now Women's and Children's Affairs), and with support from UNICEF and the ECCE Taskforce. The Government of Ethiopia formulated the framework with four main pillars, which entail (1) parental education, (2) health and early stimulation programmes (prenatal to 3+ years), (3) preschools with community-based kindergartens (4 to 6+ years), and (4) community-based non-formal school readiness programmes.[18] The third pillar calls for the establishment of a variety of preschools: in addition to private schools (kindergartens), it includes preschools in community centres, religious institutions, alternative basic education centres, and primary school compounds, which are the low-cost models of service delivery supported by a flexible arrangement between communities and governments (MoE, 2010b; Orkin et al., 2012).

[18] The first two pillars on parental education and a programme of early child health and stimulation (from the prenatal period to age three) fall under the Ministry of Health; the latter two pillars on preschools and a programme of non-formal school readiness fall under the Ministry of Education.

Guided by new ECCE Policy Framework and ESDP IV (2010-2011 to 2014-2015), the MOE has promoted the expansion of new programme type called a 'pre-primary class' or 'O-Class', a reception year for 6-year-olds before they enter Grade 1. 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 government structures. In the first year of introduction, O-Class provided one million children with immediate access to early learning, nearly three times as many children as were enrolled in kindergarten centres the year before (Woodhead, Rossiter, Dawes, & Pankhurst, 2017). The increase in O-Classes has also shifted access in ways that strengthen equity, given the rural location of the majority of O-Classes. Over the six years from 2010-2011 to 2016-2017, the gross enrolment ratio (GER) for all 4- to 6-year-olds increased from 5 percent to 46 percent. As the main driver of the rapid expansion of pre-primary coverage, O-Class is currently serving more than 2.6 million children in Ethiopia (Figure 2.2).

Figure 2.2. Enrolment Shares by Modality, Pre-Primary, 2008-2009 to 2016-2017



Note: Kindergarten (blue)—a three-year programme for 4- to 6-year olds, generally delivered in a stand-alone institution and mostly run by the private or NGO sectors; O-Class (yellow)—a one-year reception class for 6-year-olds provided in the public primary school compound; and Child-to-Child (green)—an informal programme to facilitate learning between young children with their older siblings or peers.
Source: Ministry of Education, EMIS, 2008-2009 to 2016-2017

With successes in pre-primary expansion since 2011-2012, the MoE set an ambitious vision for early learning during the planning of ESDP V (2015-2016 to 2019-2020). ESDP V, which first stated that the direct involvement of the government in ECE is a policy priority, outlined clear targets for achieving 80 percent enrolment for 4- to 6-year olds and universal pre-primary education for 6-year-olds by 2020 (MoE, 2015). It also stressed equitable access to early learning that 'quality, targeted, ECCE provision will be used as a tool to increase equity in the education system. (…) By focusing ECCE expansion first in the areas with lower educational attainment (…), the government will seek to improve the performance of children who can benefit the most from the support to transition more successfully into Grade 1' (p. 77). It signalled a significant policy shift from a limited government role in oversight of ECCE services delivered by non-state actors to the government taking a leading position in the full provision of accessible and affordable pre-primary education.

Nevertheless, with a massive influx of young children into the education system, it is inevitable that substantial challenges will arise to providing equitable access to 'quality' services in Ethiopia (Teferra & Hagos, 2016; Rossiter et al., 2018; Woodhead et al., 2017). In the next section, the landscape of early childhood education before and after the reform is illustrated, focusing on the access, programme type, equity, and quality aspects of preschool service delivery. With respect to the main interest of the present study, it helps to understand the ECE settings 'before 2010', when the EGRA 2010 cohort attended preschool, and 'after 2010', when the EGRA 2016 cohort attended preschool.

2.4.3 Landscape of Ethiopian Early Childhood Education: Before and after Reform

Early childhood education before the reform in 2010: Access, programme type, and quality. Before the early learning reform in 2010, preschool services were delivered through kindergartens operated by private institutions, NGOs, communities, and faith-based organizations. Orkin et al. (2012) identified four types of ECE providers in Ethiopia, which catered to only about 5 percent of young children over two decades. First, private preschools required a fee and were mostly located in urban areas. Second, public preschools were fee-paying but also funded by the government and were located largely in urban areas. Since private preschools and public preschools are very similar in terms of quality and both serve upper- to middle-class families, these two can be grouped into one category (Orkin et al., 2012). The third type consists of government preschools, established and operated by the *kebele* (the smallest administrative unit, similar to the neighbourhood), with teachers' salaries usually paid for by fees from parents or community contributions. These schools were located mostly in urban areas, with a few classes in rural areas.

The fourth type was the community preschool run by an NGO or religious organization. NGO schools tended to be located in urban areas, to be of relatively high quality, and to be offered to poor communities with low or no fees. Religion- or faith-based schools provided relatively low-fee education, were attached to formal religious primary schools, and served the needs of poorer parents in urban areas (Orkin et al., 2012). The religion-based preschool aimed to teach children basic literacy so they could read the Bible or Koran and instructed them in aspects of religious faith. Although there is no systematic documentation of the quality of religion-based provision, Woodhead et al. (2009) observed that, in informal faith-based preschools, children were typically taught by a single teacher, who hosted them in his own home or in the open air. [19] Many instructors had no formal training except for religious education, and the preschools did not have a set curriculum in most of the cases observed (Woodhead et al., 2009). Currently, there are few statistics and little documentation of religion-based preschools. During my interview, university-based education experts expressed that many of these institutions had declined over the years as part of an overall decline in religious primary and high schools in Ethiopia.

Early childhood education after the reform, from 2010 to 2015: Access and programme type. The Government of Ethiopia's rationale for involvement in the pre-primary sector is to offer a cost-effective model for improving children's school readiness, which can be used as a tool to increase equity in the education system. The government focuses in particular on areas with lower educational attainment and children who can benefit most from support to transition more successfully into Grade 1 (MoE, 2015). Since 2010, preschool programme types (modalities) have been diversified into two formal kindergarten programmes, KG and O-Class, and two informal programmes, CtoC and the Accelerated School Readiness (ASR) programme. Table 2.1 summarises the characteristics of the four modalities available to deliver pre-primary education in Ethiopia.

The O-Class modality currently covers about three-quarters of pre-primary service provision. O-Class is perceived as a low-cost model that is feasible to implement consistently across regions and is supported by communities and local governments (Rossiter et al., 2018). Meanwhile, scaled-up O-Class efforts did not undermine the existing systems of private kindergarten, which still play a leading role in urban centres. Kindergartens had taken up about

[19] Hoot et al. (2004) reported that mission or church pre-primary schools are generally perceived as providing good-quality education.

5 percent to 7 percent of the pre-primary enrolment in the previous two decades, although its significance faded in the overall pre-primary sector.[20] The supply of informal CtoC programmes decreased from 7 percent in 2014-2015 to 3 percent in 2017-2018. In the Tigray region, where CtoC was provided to more than half of 4- to 6-year-olds, schools had difficulty managing the informal programme regularly. The quality of instruction had been diluted during the scaled-up process due to insufficient training and supervision for older students (personal interview).[21]

Table 2.1. Four Programme Types (modalities) of Pre-Primary Education in Ethiopia

	Formal Program		**Informal Program**	
	Kindergarten	O-Class	Child-to-Child	Accelerated School Readiness
Target age group	4-6 years	6 years	4-6 years	6-7 years
Duration	Up to three years	One year (9 months)	Up to three years (part-time)	Two months (school breaks)
Main implementer	Private sector	Government	UNICEF & Government	UNICEF & Government
Main funding source	Private fees	Government & community contribution	UNICEF & Government	UNICEF & Government
Workforce	Private teachers	O-Class teachers or facilitators	Older children and CtoC trainers	O-Class or Grade 1 teachers

Source: Adapted from *Journeys to Scale* (UNICEF, 2016)

The government sectoral plan ESDP V (2015-2016 to 2019-2020) permitted a mix of modalities to be used to reach access targets (Table 2.2), stating that 'in the first years of ESDP V different approaches will be piloted and lessons learned will be used to inform expansion choice' (MoE, 2015). As an example, in 2015-2016, the MoE and UNICEF piloted the Accelerated School Readiness programmes for 6- or 7-year-old children entering Grade 1 who had not yet attended preschool. This programme provided two-month sessions (150 hours), with a focus on imparting pre-literacy, pre-numeracy, and social skills. This interim programme was expected to bridge the transition period to universal pre-primary education, given that formal ECE programmes had not yet been extended to the poorest areas (UNICEF, 2016).

[20] Nationally, kindergarten now accounts for 16 percent of pre-primary education provision in 2016-2017, down from 100 percent in 2009-2010, and its portion continues to decrease among the four modalities.
[21] Based on a personal interview with school principals in Tigray and UNICEF ECCE experts.

Table 2.2. Government's Target for Pre-Primary Education (2015-2016 to 2019-2020)

Strategies	Indicators
Access	Expand O-Class and kindergarten provision so that all children have access to at least one year of classroom-based pre-primary education
Access	Expand access to CtoC and Accelerated School Readiness programmes
Quality	Improved teaching and leadership skills in all institutions, matched with greater motivation and job satisfaction
Quality	Providing services and resources to schools to improve the physical facilities and foster a safe and healthy environment
Targets	**Indicators**
Access	Percentage of students that receive at least one year of pre-primary education will reach 100%
Access	GER for pre-primary (age 4-6 years) will rise from 34% to 80%
Access	National strategy for non-formal Accelerated School Readiness and CtoC programmes exists
Quality	Percentage of pre-primary teachers who are qualified with ECCE multi-year diploma
Quality	Percentage of pre-primary schools with qualified leader (diploma) will reach 100%
Quality	Percentage of pre-primary schools met and well above the [inspection] standards will reach 60%

Source: Compiled from ESDP V (MoE, 2015)

***Equity*.** Despite the rapid and consistent growth of preschool access, 'equity' concerns emerged. Analysis of national aggregate statistics revealed that substantial equity gaps existed in various dimensions related to location, gender, and age (Rossiter et al., 2018). First, O-Class coverage was greater in better-resourced regions. Nine established regions (including two city administrations) reached more than 50 percent of GER by 2015-2016, whereas three 'emerging' regions (Afar, Somali, and Benishangul-Gumuz) experienced slower progress, reaching 10 percent to 30 percent of GER during the same period.

Second, the equity gap appeared not only *between* but *within* regions. With the government's vision of having an O-Class in each government school, the share of primary schools with an O-Class had risen to 74 percent of schools by 2014-2015—up from 60 percent in 2011-2012 (Rossiter et al., 2018). Nevertheless, only one-third of schools had an O-Class in the three emerging regions (Afar, Gambella, and Somali), whereas in 2016-2017, 80 percent of schools were offering O-Class in all other regions. There was also a positive correlation between the 'Grade 2 to Grade 1 enrolment ratio' and O-Class coverage, which implied an unequal pattern of O-Class expansion; better managed and more affluent schools may have introduced O-Class earlier (Rossiter et al., 2018).[22] Despite the policy goal of reaching the most disadvantaged

[22] GEQIP-E has chosen to use the 'Grade 2 to Grade 1 enrolment ratio' as a key performance indicator which is a 'holistic indicator that can capture dropout, repetition, and readmission by estimating those who are lost in transition between Grade 1 and Grade 2' (World Bank, 2017). This indicator reflects aspects of educational

groups—children at risk of exclusion, dropout, and under-achievement who tend to need the service the most—had the lowest rates of preschool participation. There also was a notable gap between coverage and enrolment rates by region: 80 percent of primary schools had O-Class, but only 30 percent of preschool-age children were enrolled in O-Class. This suggested the limited capacity of the existing O-Classes, which were provided mainly in a single classroom within the school compound, and limited awareness of early learning that hindered local demands.

Third, there was a gender gap in preschool access that favoured boys over girls, although it was less severe than the regional variation. The Gender Parity Index widened slightly from 0.98 to 0.95 on average between 2010-2011 and 2015-2016, yet it should be noted that the gender gaps were more pronounced in the three 'emerging' regions: Somali (0.84), Gambella (0.89), and Benishangul-Gumuz (0.90) (Rossiter et al., 2018). The fourth dimension looked at the age distribution of O-Class enrolees. The average share of 6-year-old children who enrolled O-Class (a target age group for O-Class) was merely 42 percent, with huge regional variation, from 20 percent in Somali to 90 percent in Tigray (Rossiter et al., 2018). While many 6-year-olds still were missing out on the pre-primary education opportunity before entering Grade 1, the multi-age composition raised concerns about the supply constraints (e.g., 'churning' in O-Class more than a year) and deterioration of quality due to the absence of age-appropriate curriculum and pedagogy.[23]

These huge regional disparities were closely linked to the Regional Education Bureaus' (REBs) varied response to the reform initiatives directed by the central government. Since the Ethiopian education system is highly decentralised and features the transfer of financial management and decision-making authority from the upper to lower levels of government (MOE, 2002), REBs carry the primary responsibility for operationalising the policy targets stipulated in the national sectoral plans. Nevertheless, there was 'incoherence between delegated national objectives and local preferences and capacities, such that O-Class access

---

attainment, advantage, and efficiency for each woreda. A lower 'Grade 2 to Grade 1 enrolment ratio' suggests that schools and woredas have lower levels of educational attainment and efficiency which would classify them as priority areas for services (Rossiter et al., 2018).

[23] The notion of 'churning' is from the early grade inefficiencies identified in emergent primary education systems (Crouch, 2015). It indicates that Grade 1 or Grade 2 enrolees tend to stay in the early grades for multiple years without any mark of grade repetition or readmission (usually marked 'new entrants' several times). Anecdotally, a similar tendency has been observed in O-Class.

[wa]s provided unequally, for a subset of the target population' (Rossiter et al., 2018, p. 19). The implementation processes of the early learning reform were far from uniform and were highly dependent on the human and financial resources of local governments within each state.

While the inherited regional historical differences have been mirrored in the different pathways to the scale-up of early learning, this often led to innovative service delivery tailored to the needs of local communities. For example, in SNNP, to minimise the travel distance to primary schools for children in remote areas, O-Classes were established not only in the regular government primary schools but in the non-formal community-based organizations within religious institutions (churches, mosques), farmer's training centres, and established *kebele* (community centres).[24] The various forms of O-Class ('satellite') were grouped into a single school ('cluster') for better management purpose. This approach led to the rapid growth of O-Class participation in SNNP, which reached 65 percent in 2016, nearly double the national average of 33 percent (MOE, 2016). In Afar, one of the 'emerging' regions with few resources, informal preschool programmes (CtoC and ASR) were further adapted to a mobile education system (Rossiter et al., 2018) in order to accommodate the needs of pastoralist communities.

Finance. Critically, the policy directive was not accompanied by any form of financial support to deliver the national-scale increase in preschool access. Even after the government target was set to reach *all* 6-year-olds (MoE, 2015), the national plans allocated no more than 3 percent of the education budget to pre-primary education (Rossiter et al., 2018).[25] Rolling out an ambitious plan of early learning reform without any financial commitment had the potential to place an additional burden on an already overextended primary education system (Orkin et al., 2012). This could have turned promising ECE initiatives into a mere downward expansion of primary education, which would have replicated many problems raised by the earlier reform in universal primary education.

[24] According to the interview with SNNP REB staff, out of 13,938 O-Class institutions, 6,013 (or 43.8%) institutions are in-school O-Classes, 5,521(or 39.6%) are O-Classes in religious institutions or farmers' training centres, 1281 (or 9.2%) are O-Classes in established community centres. The rest is child-to-child (205, or 1.5%) and kindergarten (828, or 5.9%).

[25] ESDP V (2015/16-2019/20) allocated 3 percent of its total budget to early learning, in addition to any resources supporting oversight via the 'Admin' budget line (Rossiter et al., 2018, p. 11). Most recently, in 2018-2019, it was reported that about 7 percent of the education budget was allocated to pre-primary education.

With a lack of sufficient financial support, the provision of O-Classes was heavily dependent on community contributions. Regions reported that communities were involved in financing, managing (through a school committee), and taking ownership of O-Classes, and more recently in supplementing teachers' salaries and the cost of materials (Rossiter et al., 2018). This implies that the rapid expansion of pre-primary education could not be detached from the community's perception and preference for early learning, but it also favoured resource-rich areas where the communities could afford to set up new O-Classes. In 2016-2017, the government introduced school grants for O-Class (a capitation grant for 6-year-olds enrolees), which represented the first explicit financial source provided to schools for pre-primary. Although REBs perceived school grants to be the major source of financing for O-Classes, they in fact had limited capacity to reach the schools least able to establish O-Classes, which could further exacerbate access inequality (Rossiter et al., 2018).

Quality. With an unsystematic approach to the preschool reform in Ethiopia, substantial challenges to providing 'quality' ECE services occurred. Several studies conducted in Ethiopia attest to the actual lack of quality and the lack of a system to yield quality outcomes. Regarding the former, Teferra and Hagos (2016) and Dowd et al., (2016) pointed out the pervasive lack of quality services in pre-primary education. Preschool, especially O-Class, suffers from many challenges, such as the shortage of trained and qualified teachers/facilitators; the lack of developmentally appropriate curriculum and learning materials; the lack of adequate infrastructure and safe school facilities; and a lack of incentives for teachers assigned to O-Class.

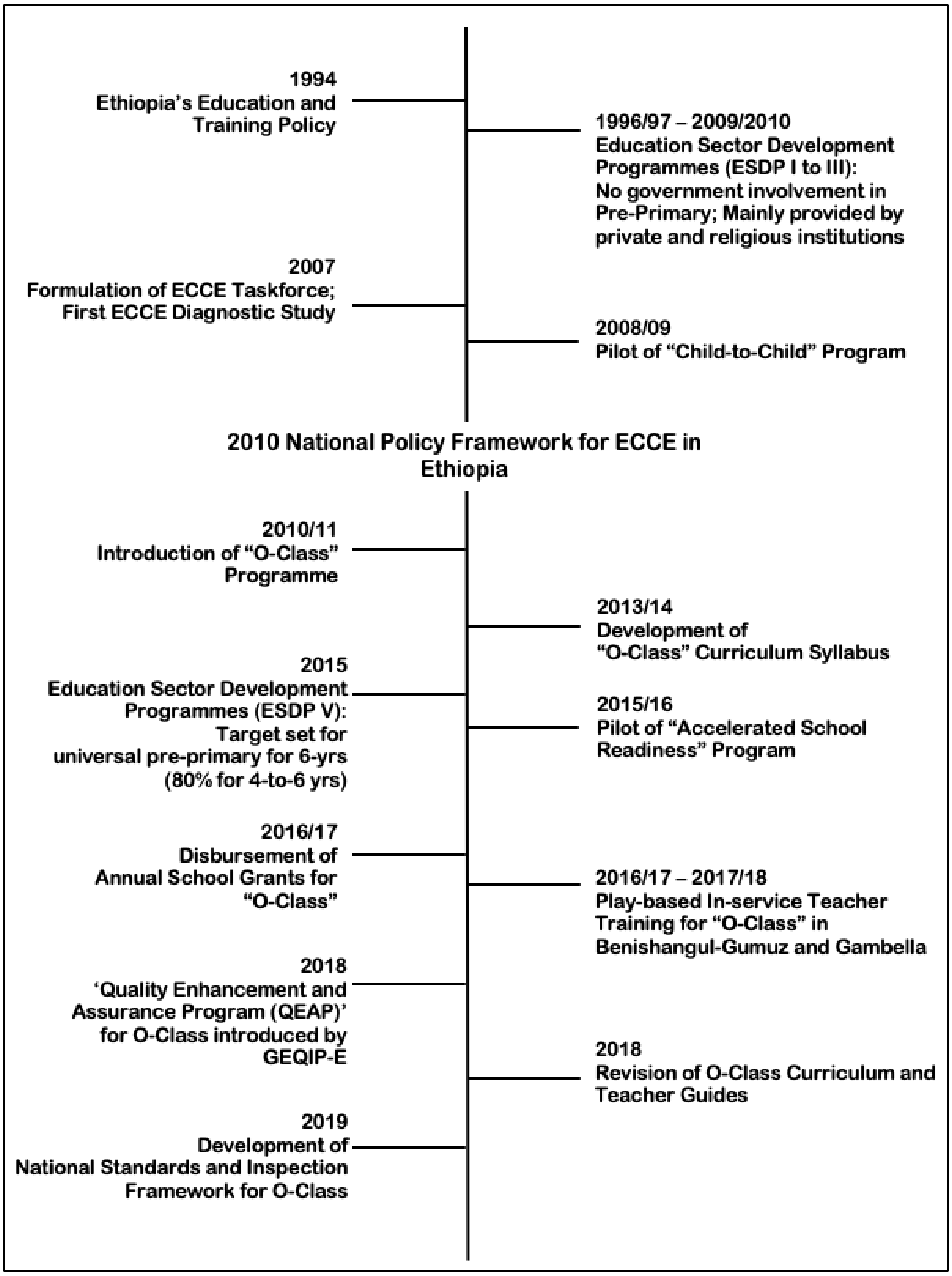
Regarding the latter, Woodhead et al. (2017) and a recent diagnostic report by Rossiter et al. (2018) have noted the lack of a system to deliver quality services and the lack of coherence across multi-tiered systems. The research has found that '*downwards* delegation from federal and regional levels to woredas and schools suffers coordination and communication weaknesses' (Rossiter et al., 2018, p. 29). Without attending to the system's readiness to promote early learning (e.g., a shortage of ECE specialised administrators at all levels of government), coherent and quality services are unlikely to become a reality. Collectively, previous research on the status of O-Class warned that low-quality, unequal preschool service provision will not deliver on the potential of ECE and could even have some detrimental consequences for child development (Teferra & Hagos, 2016; Woodhead et al., 2017). As

Kagan et al. (2015) noted, this is consistent with the global understanding of the scale-up process that ‘programmatic expansions that simply build on a dysfunctional structure are unlikely to produce significant gains for quality, much less for the equitable distribution of services or for their sustainability’ (p. 9).

New initiative for quality improvement of O-Class: 2016 to present. As the early learning system in Ethiopia is constantly evolving, the government’s focus is gradually shifting to improving quality. Figure 2.3 summarises a timeline of key developments in the establishment and rollout of pre-primary education in Ethiopia. With respect to the latest initiatives, from 2016-2017 to 2017-2018, a region-wide programme of in-service teacher training for O-Class teachers was implemented in two historically disadvantaged regions in Ethiopia—Benishangul-Gumuz and Gambella (World Bank, 2017). This entailed the development of child-friendly, play-based curriculum and teacher guides in local languages, as well as an intensive one-month teacher training for all O-Class teachers in two regions.

In 2018, the General Education Quality Improvement Programme for Equity (GEQIP-E)—a third-phase of the government’s education reform package—formally introduced a Quality Enhancement and Assurance Programme (QEAP) for O-Class (World Bank, 2017). This comprised two key components—quality enhancement (QE) and quality assurance (QA). The QE component, which aims to improve pedagogical practices in the classroom, includes teacher preparation and professional development, curriculum and teaching and learning materials for O-Class, training for management and supervision, and an orientation programme on early learning for parents and communities. The QA component, which aims to create a quality assurance mechanism, entails the establishment of national standards, school inspection for O-Class, and quality EMIS data collection for pre-primary. Overall, QEAP is expected to provide a comprehensive package of interventions to systemically improve the quality of O-Class within a coherent framework of support. Interrelated activities of the QEAP are currently being implemented across regions in Ethiopia.

Figure 2.3. Timeline of Key Developments of Pre-Primary Education in Ethiopia



2.5 The Present Study

In the present study, I aim to build knowledge of the influence of early learning reform in Sub-Saharan Africa by focusing on the country experiencing the most rapid expansion of preschool access: Ethiopia. Specifically, I aim to examine the change in the patterns of association between preschool attendance and students' learning outcomes before and after a large-scale expansion of public preschool (O-Class). Using the EGRA, which was administered in 2010 and 2016 to a regionally representative sample of Ethiopian students, this study addresses two primary research questions, as follows:

1. What is the difference in the test score *distribution* of second- and third-grade students' reading achievement, as measured by oral reading fluency (ORF), between preschool attendees and non-attendees before and after the early learning reform?
2. Does the early learning reform (i.e., large-scale expansion of preschool) *strengthen* or *weaken* the role of preschool attendance in predicting second- and third-grade students' reading achievement, as measured by (1) EGRA test scores, and (2) the probability of being a non-reader or a proficient reader?

This study first illustrates the trends in preschool participation between 2010 and 2016 and examines the determinants of preschool attendance in each period. The aim is to capture how pre-primary education in Ethiopia shifted from the elite to the mass system, and to what extent this induced compositional changes between children who attended preschool and those who did not. Using the novel approach of estimating the achievement gap over time, I address my first research question on the difference in test score *distribution* between preschool attendees and non-attendees during the early learning reform.

My second research question aims to determine whether preschool attendance is predictive of students' early grade reading achievement and how this predictive role has evolved during the large-scale expansion of preschool. The following outcome indicators are used to address this question: (1) EGRA test scores from six sub-tasks, such as ORF and reading comprehension, and (2) the probability of being a non-reader or a proficient reader. While the test scores provide useful information on students' average performance, the latter two outcome indicators on the probability have certain advantages. First, they are less sensitive to language characteristics that vary by region. This is particularly relevant to the multilingual culture in Ethiopia, where the most progressive, longstanding mother tongue instruction policy was adopted in 1994 (Seid, 2016). The Government of Ethiopia currently uses 51 official mother tongues for teaching and textbooks in Grade 1 to Grade 4. Second, these measures can provide more comparable metrics to analyse the trends over time than average test scores, which is the main interest of the current study. Third, these measures have better interpretability. Initially, the reading proficiency levels were developed for better communication with government officials (Gove & Wetterberg, 2011). The results based on reading proficiency levels could more explicitly articulate incremental expectations for student performance in early grade reading and were comparable to the Minimum Learning Competencies in the national curriculum (Piper, 2010; RTI, 2015).

A strength of the current study is the ability to track the relation between preschool attendance and students' academic achievement over the reform period. The study also focuses on the early grades in primary schooling, from a reception year through Grade 3. Early grades are focal years in which policymakers and practitioners pay special attention to equipping children with fundamental skills for lifelong learning. In LMICs in particular, intervention in the early grades is instrumental in preventing the high dropout rate and grade repetition that have burdened the efficiency and effectiveness of the education system (Bashir, Lockheed, Ninan, & Tan, 2018; Crouch & Merseth, 2017).

2.6 Data and Sample

Data used for this study come from the Early Grade Reading Assessment, an influential tool used to assess students' early academic ability in reading acquisition. The EGRA is primarily designed to collect information on individual-level early literacy skills, as measured by a variety of sub-tasks from letter recognition to reading comprehension. EGRA was introduced in Ethiopia in 2010, with the aim of providing a national-level diagnostic of students' reading levels within the rapidly changing primary school environment (Piper, 2010). As a school-based assessment, EGRA is administered one-on-one to students currently enrolled in Grades 2 and 3.[26] The EGRA dataset contains basic information on individual-, family-, and school-

[26] Grade 2 and Grade 3 were selected instead of Grade 1, as they represent students' abilities after some years of schooling (AIR, 2016).

level characteristics and, crucially for my purpose, it asks students whether they ever attended preschool before entering primary school.

Broadly speaking, the motivation for creating the EGRA was to ensure timely access to information that could inform learning improvement efforts in low-income countries (Dubeck & Gove, 2015). Drawing from an extensive body of research on early reading acquisition, including a U.S.-based instrument, the Dynamic Indicators of Basic Early Literacy Skills (Good & Kaminski, 2002), two main principles underpin the EGRA (Gove & Wetterberg, 2011). First, provided that 'acquiring reading skills is a multi-phased process', EGRA breaks down each assessment into sub-tasks that correspond to the building blocks of reading acquisition. Second, with the knowledge that 'learning to read is likely to vary by language and context', EGRA uses a method-independent approach focused on core foundation skills. In other words, the EGRA instrument does not reflect any particular method of reading instruction, but instead allows the assessment design to be flexible, based on the linguistic context.[27]

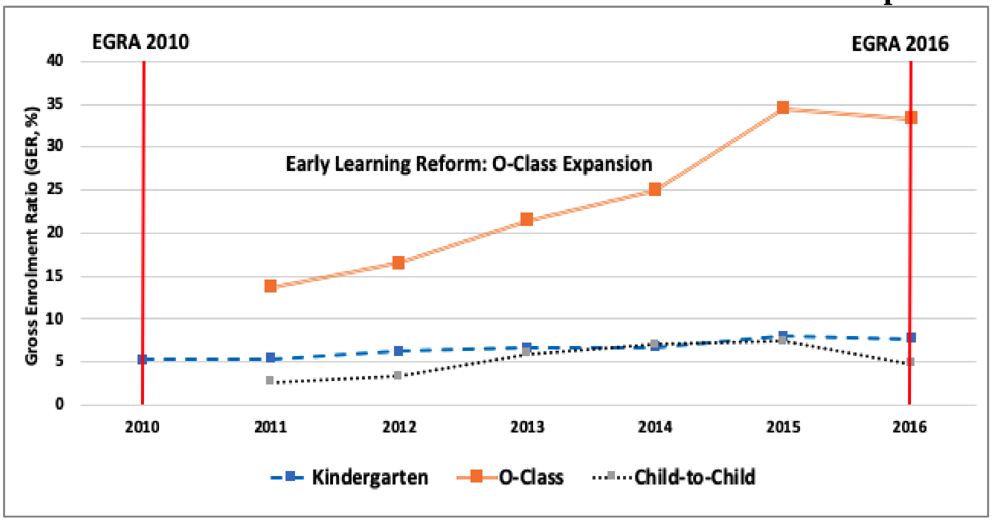
EGRA, which was developed in 2006 by multiple international agencies, including USAID, the World Bank, and RTI International, has been adapted for use in more than 70 countries and more than 120 languages.[28] The EGRA has served multiple purposes: as a baseline measure of early reading acquisition (UNESCO, 2014); as a guide for the content included in an instructional programme (Gove & Cvelich, 2011); and as an impact evaluation tool for a literacy or educational intervention to inform the transition from pilot to country-wide scale-up (Piper et al., 2018; Piper et al., 2014). Despite some limitations of the instrument itself, such as containing a limited construct for measuring early literacy or not being intended for use as a high-stakes accountability measure (see Bartlett, Dowd, & Jonason, 2015), the EGRA's clear theoretical framework and consistent application procedures provide valid and reliable information for each of the purposes listed above (Dubeck & Gove, 2015).

An important advantage of using the EGRA for the present study, in addition to demonstrating a link between preschool attendance and early literacy skills, is that the two assessments straddle the period of early learning reform in Ethiopia. In particular, this study leveraged data

[27] In Ethiopia, the instrument has been adapted locally. In order to test the reliability and validity of the various sub-tasks in six languages, all assessments were piloted extensively in each region prior to use (Piper, 2010).
[28] For updated figures (international), see the EGRA tracker at www.eddataglobal.org.

from two regionally/linguistically representative samples of Grade 2 and Grade 3 students that were collected in 2010 (pre-reform) and 2016 (post-reform). These datasets provide a unique opportunity to measure changes in the contribution of attending preschool during the transition to the mass system, concomitant with a sudden influx of children previously excluded from early childhood education. Figure 2.4 presents the enrolment trends in pre-primary education and the EGRA administration in Ethiopia. The survey instrument and data collection procedures for EGRA 2016 were modelled after the 2010 EGRA administration in a comparable manner, and thus were selected for this study.[29] Three additional rounds of the EGRA were administered in 2013 and 2014; however, these were not included in the present study, either because no full dataset was available from the agencies, or because it was conducted as a pilot only for newly introduced languages.[30]

Figure 2.4. EGRA Administration before and after the Reform in Ethiopia



Source: MOE Education Annual Abstract Statistics 2010-2011 to 2016-2017 G.C., MoE, Ethiopia

---

[29] EGRA provides cross-sectional information as a new sample of schools and students were drawn in each round. The EGRA Ethiopia 2010 and 2016 datasets were obtained through the USAID Reading Network and USAID Ethiopia, respectively.

[30] In total, three agencies have been involved in Ethiopia's EGRA administration: 2010 and 2014b EGRA by Research Triangle Institute (RTI); 2013 and 2014a EGRA by fhi360 as part of Improving the Quality of Primary Education Programme (IQPEP); and 2016 and 2018 EGRA by American Institute of Research (AIR) as part of USAID Reading for Ethiopia's Achievement Developed Monitoring and Evaluation (READ M&E). It is inevitable to drop 2013 and 2014a EGRA, since only partial information from students' responses is available, due to miscoded items. I've contacted the agency (fhi360), but due to the absence of the experts who were involved, full information is not available. Administration of EGRA 2018 was completed by AIR in June 2018 and data were to be released in early 2019.

2.6.1 Constructing a Comparable Dataset between EGRA 2010 and EGRA 2016

1) The scope of the sample

Regions and languages. The original EGRA 2010 and EGRA 2016 samples consisted of 13,079 students (from 8 regions, speaking 6 languages) and 12,124 students (from 5 regions, speaking 7 languages), respectively. Approximately 90 percent of the population in Ethiopia speaks at least one of the languages spoken by the students who were assessed (AIR, 2016). For comparability between the two cohorts, the present study used final samples of 9,121 and 8,332 students from each round, which were limited to five regions: Tigray, Amhara, Oromia, Somali, and the Southern Nations, Nationalities, and Peoples' region (SNNP), as shown in Table 2.3. This was a crucial step to ensure that differences in patterns across early literacy achievement were not driven by regional differences in terms of who participated in the EGRA administration.

Table 2.3. Comparable Sample between EGRA 2010 and EGRA 2016

Region	Language	**EGRA 2010**	**EGRA 2016**
		Number of students	Number of students
Tigray	Tigrigna	1,537	1,709
Amhara	Amharic	2,259	1,748
Oromia	Afan Oromo	2,442	1,749
Somali	Aff Somali	1,163	1,352
SNNP	Sidamu Afoo	1,720	1,774
Total		9,121	8,332

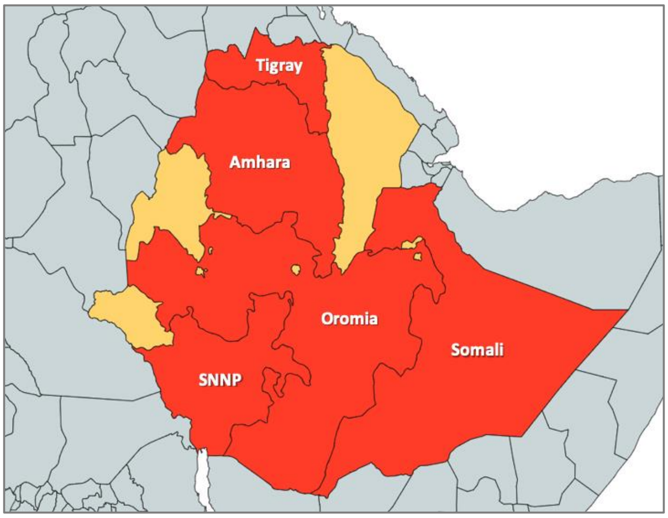
Source: EGRA Dataset 2010, 2016, USAID

Importantly, as shown in the map in Figure 2.5, these five selected regions are where 94 percent of Ethiopia's 4- to 6-year-old children live (Rossiter et al., 2018).[31] These regions reflect a significant amount of linguistic and cultural diversity. The sample in four regions—Tigray, Amhara, Oromia, and Somali—are regionally representative, whereas the sample in SNNP is considered to be language representative, for the population speaks *Sidamu* languages. In SNNP, out of the 32 mother tongues officially adopted as a medium for textbooks and instruction, Sidamu is the most predominantly spoken and was included in both EGRA 2010

[31] All population statistics are estimates, projected from a national census conducted in 2007. The distribution of the 4- to 6-year-old population represents the distribution of total population (any age) across regions (Rossiter et al., 2018, p. 8).

and EGRA 2016.[32] The final sample excluded two urban city centres (Addis Ababa, Harari), where private kindergartens are more prevalent, one small region (Benishangul-Gumuz), and two local languages (Hadiyya and Wolayitta), all of which were administered only once in either EGRA 2010 or EGRA 2016.[33]

Figure 2.5. Map of Ethiopia: Five Sample Regions



Sampling frame and weights. To ensure regional/language representativeness, all analyses in this study were estimated using survey weights provided by each round (AIR, 2016; Piper, 2010). In the sampling frame of EGRA 2010 and EGRA 2016, once the regions, woredas, and schools were randomly selected, the stratified sample by gender and grades from each school participated in the assessment. EGRA 2010 specifically applied a three-stage stratified sampling framework by using a proportional population sampling at the regional and school levels, and systematic sampling at the classroom level (Piper, 2010). EGRA 2016 focused on ensuring that a certain number of schools and students (sample size) were randomly selected in each region to obtain the desired statistical power for regional/language representativeness (AIR, 2016).[34] In both rounds, about 40 students were randomly selected from each school,

[32] The 2007 census reported that the predominantly spoken mother tongues in SNNP include Sidama (19.59%), Welayta (10.48%), Hadiyya (8%), Gurage (7.13%), Gamo (6.9%), Kafa (5.36%), and Amharic (4.10%). EGRA 2014 (pilot) and EGRA 2016 included two additional languages (Hadiyya and Wolayitta) from the SNNP region.
[33] For example, the gross enrolment ratio of kindergarten in Addis Ababa was 87.5 percent in 2010-2011 and 90.8 percent in 2016-2017, according to the National Education statistics.
[34] For the analysis of EGRA 2010, the Stata ‘svy’ command was used to establish the parameters for each level of selection; for EGRA 2016, proportional survey weights were used to adjust a fixed number of students from each school to actual school-level enrolment in a particular grade. With respect to the relation of sampling with other reading project, EGRA 2010 provides a baseline for the IQPEP and sampled the IQPEP schools purposively,

with equal proportions of girls and boys, and students from Grade 2 and Grade 3. The sample sizes in both EGRA 2010 and EGRA 2016 were larger than expected to test statistically significant differences (AIR, 2016; Piper, 2010).

It should be noted that no EGRA administration achieved a complete random sampling, due to security reasons such as conflicts among ethnic groups, internal migration, and natural disasters (e.g., droughts, floods). In EGRA 2010, due to the political instability of the Somali region during the election period, convenience sampling was done in that region at the woreda level and random sampling at the school level. Similarly, EGRA 2016 avoided areas listed as the top priority zones by UNICEF's Emergency Education Cluster report, which were in all five regions. To avoid any loss of sample, replacement schools were randomly selected from the initial sampling stage. About one-fifth of the initial sample schools were replaced during the data collection, due to security or other logistic issues (e.g., school closure). The largest replacement in 2016 was made in Oromia, followed by Amhara, Somali, Tigray, and SNNP (Sidamu).

Lastly, from the selected regions and language groups, non-response rates were very low in both EGRA administrations. In EGRA 2010, item non-response was almost none for outcome variables and preschool attendance (less than 0.05% for both) and for a set of control variables (less than 0.01%). Similarly, in EGRA 2016 there were no missing values in outcome variables and preschool attendance. For the control variables collected by the self-reported survey, the non-response rate was very low, from 0.003 percent in language of instruction to 0.02 percent in father's literacy. Considering the very low item non-response, I used listwise deletion (also known as complete case analysis), which is less likely to introduce bias if the data are 'missing at random' and provide accurate estimates of true standard errors (Allison, 2002).

2) EGRA instruments and administration methods

Since the EGRA created a locally adapted instrument in 2010, subsequent EGRA administration in 2013, 2014, and 2016 consistently applied test items comparable to the initial

---

along with RTI schools, yet it doesn't affect the representativeness of the sample (Piper, 2010). Although EGRA 2016 was administered in support with the READ M&E project, this project was delivered at a national scale without phasing to allow for a control group (AIR, 2016; Gove et al., 2017). Thus, EGRA 2016 cannot be taken as a measurement of impact, nor can it be directly tied to implementation.

EGRA instrument. While retaining international comparability, the EGRA instrument appears to fit well with the expected learning competencies set by the Ethiopian curriculum for Grade 2 and Grade 3 students (Piper, 2010). Of the seven EGRA sub-tasks, the one on phonological awareness was excluded from the present study, due to comparability issues between the two tests. While EGRA 2010 used three different phonological awareness tests—initial letter sounds, final letter sounds, and the number of sounds—differentiated by language characteristics, EGRA 2016 administered only an initial letter sounds test, regardless of language or region.[35]

There was a partial revision of the original instrument in 2016, based on the newly developed mother tongue curriculum in Ethiopia. To establish item comparability between the 2010-2014[36] and 2016 ORF sub-tasks, the survey was conducted with an additional sample of 1,400 students during the EGRA data collection (AIR, 2016).[37] By employing the *common-person* research design, the same students took part in more than one version of the assessment to detect any differences that could be attributed to the instrument characteristics, rather than to student characteristics. Based on the findings of this survey, the test items from the 2010-2014 and 2016 EGRA instruments were equated, and it verified the comparable use of two different administration methods—paper-based versus tablet-based (*tangerine*) EGRA.

2.6.2 Model Variables

1) Key explanatory variable

Preschool attendance. Preschool attendance was measured retrospectively through the student questionnaire. The following question was asked on the 2010 EGRA: 'Did you go to pre-primary or kindergarten?' and the following on the 2016 EGRA: 'Did you go to a nursery or preschool (zero-class) before Grade 1?'[38] I categorised students as having attended 'preschool'

---

[35] Initial letter sounds for SNNP and Harari; final letter sounds for Tigray, Oromia, Somali, and Harari; and number of sounds for Amhara, Benishangul-Gumuz, Harari, and Addis Ababa (multiple tests administered in one region).
[36] The same test items were used between 2010 and 2014.
[37] Besides the 12,124 original EGRA 2016 sample, an additional 1,400 students (200 students in five schools per language) were sampled and participated in the pilot survey using *common-person* approach.
[38] 'Nursery' in Ethiopia is informal childcare provision by the private sector. Qualitative field research by Orkin et al. (2012) describes that nursery could be either 1- or 2-month childcare service or a half-semester programme that combined childcare and education. Nevertheless, provided that EGRA was implemented exclusively in public primary school, there was little chance that EGRA sample children attended nursery attached to private school.

if they attended any of these programmes, regardless of service provider or institution. 'Preschool attendance' is thus defined as including a broad set of centre-based or classroom-based ECE experience in formal and informal, public and private institutions.[39] This is one of the major limitations of the present study, given that information collected by these questions cannot provide any details on the pre-primary service provision they received, such as preschool type, length or duration, and quality of instruction. There also is a possibility for biased recall of programme participation, as students were asked about preschool attendance retrospectively. The EGRA dataset, however, did not allow researchers to conduct any critical assessment of whether children in Grade 2 and Grade 3 reported correctly.

Instead, to check the reliability of self-reported 'preschool attendance' in the EGRA, I looked at the trends of enrolment in pre-primary education reported by the official education statistics in Ethiopia (referred as the Education Management and Information System, EMIS hereafter). Table 2.4 displays the trends over academic years 2007-2008, 2008-2009, 2013-2014, and 2014-2015, when the EGRA 2010 and EGRA 2016 cohorts belonged to the preschool-eligible group. On average, between 2007 and 2014, the gross enrolment rate soared from 4 percent to 41.3 percent in the national statistics, while pre-primary participation of the EGRA sample increased from 14.2 percent to 38 percent.[40] Some discrepancy in earlier years (e.g., 4% in EMIS and 14% in EGRA) may stem from underreporting issues in the kindergarten programmes run by the private sector, NGOs, and faith-based organizations (MoE, 2010, 2012). In terms of trends by different ECE programme types, the enrolment rate in O-Class increased rapidly, from none to 24.4 percent between 2007 and 2014, which has continuously provided more than a half share of the ECE programmes. In contrast, the enrolment rate in kindergarten was starkly stable at 4 percent over this period. Two major data sources captured similar trends in pre-primary enrolment. This sheds light the unprecedented public attention to pre-primary education in recent years, mainly through a massive expansion of O-Class rather than an increase in private kindergartens or informal programmes.

To my knowledge, there are no official statistics for nursery. Since the inclusion of nursery in the 2016 questionnaire was inappropriate, I asked the administration agency to exclude this choice. For future EGRA administration, one additional question will be asked if a student responds to 'yes' to preschool attendance: which type of preschool they attended.

[39] A broad preschool category has been used in studies in Sub-Saharan Africa (Bietenbeck et al., 2017, for Kenya, Tanzania; Hungi & Ngware, 2018, for Uganda).

[40] The figure from the national statistics is the average between 2007-2008 and 2008-2009, and between 2013-2014 and 2014-2015; the figure from EGRA is weighted value.

Table 2.4. Participation in Pre-Primary Education, EMIS and EGRA (5 regions)

	EMIS 2007/08	EMIS 2008/09	EGRA 2010	EMIS 2013/14				EMIS 2014/15				EGRA 2016
	KG	KG		O-Class	KG	CtoC	Total	O-Class	KG	CtoC	Total	
Tigray	1	10	12.6	9.6	6.2	64	79.8	28	7	64	99	50.0
Amhara	2.0	2.2	6.1	31.3	2.6	8	41.9	31	2	8	41	22.0
Oromia	3.0	3.4	21.9	12.4	5.4	2	19.8	17	6	2	24	23.6
Somali	0.6	0.6	10.0	0.6	3.5	0	4.1	0.6	4.7	0.1	5.4	27.2
SNNP	3.0	3.5	18.2	35.5	5.1	6	46.6	46	5	6	57	64.4
Total	3.9	4.2	14.2	17.9	4.56	16	38.4	24.4	4	16	44.2	37.9

Note: (1) for EGRA, weighted values are presented; (2) EGRA 2010 sample students were of preschool age (4- to 6-year-olds) between 2007-2008 and 2008-2009; (3) EGRA 2016 sample students were of preschool age between 2013-2014 and 2014-2015, after the early learning reform was initiated in 2010; (4) EMIS stands for Education Management and Information System; (5) KG and CtoC stands for Kindergarten and Child-to-Child Program, respectively; (6) for Somali, data on EMIS 2014-2015 is replaced by data from EMIS 2015-2016, considering the data availability; (7) SNNP in EGRA includes only the Sidamu language.
Source: MoE Education Statistics Annual Abstract, 2007-2008, 2008-2009, 2013-2014, 2014-2015 E.C.; EGRA Dataset 2010, 2016, USAID.

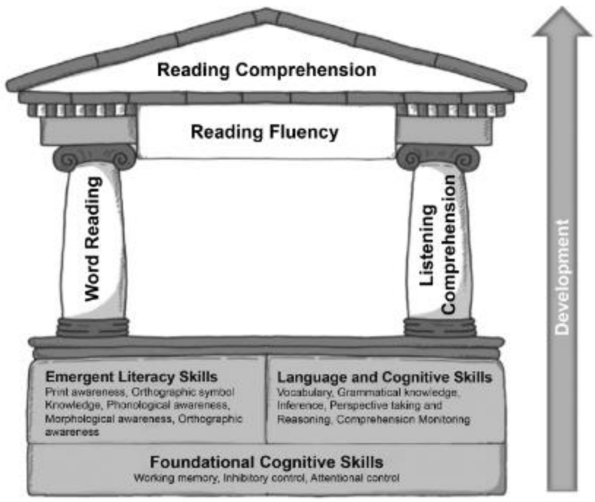
Although the national figure helps us grasp the overall trends, the direct comparison between EMIS and EGRA data calls for some caution, for the following reasons. First, the gross enrolment ratio was calculated based on the age 4-6 population, yet O-Class explicitly targets 6-year-old children. Moreover, this population estimate was drawn from the 2007 census, with a possible error in population estimation. Second, the EGRA represents one language group (Sidamu) from the SNNP region, which is about one-fifth of the entire region’s population, whereas EMIS represents the average pre-primary enrolment in the SNNP region. Third, data from the Somali region have not been properly reported in the annual statistics, due to the frequent school closures because of political instability and natural disasters, as well as weak data management capacity in the region. Fourth, the inclusion of the CtoC modality in the national statistics adds complexity in analysing the trends. For example, CtoC, which is dominantly provided in Tigray, is a rather informal programme for 4- to 6-years-old, where older children (Grade 5/6 students) teach younger children under the guidance of adults. Due to the ambiguity of the question being asked in EGRA 2016, it is not clear whether respondents count CtoC as their early learning experience before primary schooling, or they attended both CtoC and O-Class.

2) Outcomes of interest

Test scores from the EGRA sub-tasks. Academic outcomes were drawn from a direct assessment of students’ early literacy skills using the EGRA, which was conducted at the end of the academic year in 2010 and in 2016. The first set of outcome variables came from test scores measured by six EGRA sub-tasks, including (1) ORF; (2) letter name recognition; (3)

familiar words recognition; (4) invented words recognition; (5) reading comprehension; and (6) listening comprehension. ORF is the indicator from the EGRA most frequently used as a valid proxy for literacy skills (Piper et al., 2014).[41] I also used other EGRA sub-tasks to consider the interconnectivity among all measures in the development of reading skills (Bartlett et al., 2015), as illustrated in Figure 2.6. To achieve ORF and reading comprehension, the ultimate goals of reading skills development, students started with the foundational cognitive skills (e.g., working memory, inhibitory control, etc.), and the emergent literacy and languages skills (e.g., phonological awareness, vocabulary, etc.) evolved through word recognition and listening comprehension.

Figure 2.6. Skills and Abilities of Literacy and Reading Development



Source: Kim, Boyle, Zulkowski, & Nakamura (2016, pg. 9)

Table 2.5 summarizes the property of each EGRA sub-task. During a 15-minute, one-on-one oral assessment, primary grade students were asked by a trained enumerator to identify letter names or letter sounds and to read aloud common words and a brief grade-level passage (AIR, 2016; Piper, 2010). The first four tasks were timed, with a one-minute limit. Students read from a printed paper sheet while the enumerator recorded their answers on a digital tablet. For the latter two tasks, students first read a passage and answered questions about the passage just read to determine reading comprehension; students then listened to a brief passage read aloud

[41] Researchers in the U.S. have found correlations between .49 and .83 for ORF scores and achievement tests (Barger, 2003; Shaw & Shaw, 2002; Silberglitt et al., 2006; Vander Meer et al., 2005; Wilson, 2005, cited from Piper et al., 2014).

by an assessor and answered the assessor's five questions to determine listening comprehension.

To ease the interpretation of the coefficients, I used both raw test scores and standardised scores (z-scores). The raw test scores are expressed as 'correct letters/words per minute' or 'percentage of correct answer'. A student's standardised score was based on the mean and standard deviation of the raw scores of the corresponding test. For a particular outcome measure (e.g., ORF), this represents how much higher or lower an individual's score was compared to other reading fluency test takers, as expressed in standard deviation units. Meanwhile, consistent with the global EGRA application (Gove & Wetterberg, 2011), each score from the six sub-tasks was used separately, instead of calculating the composite score.

Table 2.5. Property of EGRA Sub-Tasks

Sub-Tasks	Property, Assessment Method
Oral reading fluency (ORF)	Assess the ability to read with speed, accuracy, and proper expression. This task examines whether students in Grades 2 and 3 were able to read aloud a passage with speed and accuracy with grade-appropriate words, as presented in the student workbook. ORF has a strong correlation with reading comprehension (Fuchs, Fuchs, Hosp, & Jenkins, 2001)
Letter name recognition (or fidel identification)	Assess knowledge of the alphabetic principle, the foundation of learning to read. The alphabetic principle is the understanding that words are composed of sounds (i.e., phonemes) and that letters (i.e., graphemes) are symbols that represent those sounds. Research in other languages has suggested that reading skills progress only after 80 percent of letters are mastered (Seymour, Aro, & Erskine, 2003)
Familiar words recognition	Assess the ability to recognise and read high-frequency words (determined by the most commonly used words in textbooks). Unlike ORF, this task is not presented as a story or complete text. From this task, we can attain a measure of decontextualised decoding skills, which is a distinct skill from reading comprehension from a text (Gove, 2009).
Invented words recognition	Assess the ability to decode one- and two-syllable non-words that could plausibly exist in the language in question. Compared to familiar word recognition, this task allows us to measure whether students can read *non-sight-read* words (Hirsh, 2013).
Reading comprehension	Assess understanding of the text in a passage and the ability of pupils to answer factual questions and make inferences based on what they read in ORF subtasks. Research indicates that the ability to correctly understand and interpret oral stimuli (linguistic comprehension) and make meaning from what is heard is a core skill related to reading comprehension (Hoover & Gough, 1986; Kamhi & Catts, 1991).
Listening comprehension	Assess some of the core dimensions of listening related to short-term memory, discriminating among distinctive sounds, detecting key ideas, and guessing meaning from context. In this task, students will answer several questions from a simple story read aloud by the administrator in an interactive situation.

Source: AIR (2016, pp. 13-15)

Non-reader and proficient reader based on Ethiopia's reading proficiency level. In addition to the EGRA test scores, two outcome variables, *non-reader* and *proficient reader,* were used in the present study. These indicators were widely used, especially for cross-country comparison, based on students' achievement on the ORF assessment (Gove & Wetterberg, 2011; Kelly & Graham, 2017). Non-reader was measured by the proportion of second- and third-grade students who could not read a single word correctly. Proficient reader was measured by the proportion of second- and third-grade students who were at the upper levels of reading proficiency. In Ethiopia, '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 students exhibiting relatively *functional* reading proficiency levels (AIR, 2016).

As shown in Table 2.6, the reading proficiency level is defined distinctively by languages and grades in terms of *correct words per minute*, as measured by ORF. The range of each benchmark was developed through intensive data-driven consultation with the MoE, regional language experts, and key stakeholders (RTI, 2015). To establish benchmarks corresponding to students' reading performance and the national curriculum, regional language experts created language-specific metrics, such as looking at the intervals of ORF scores achieved by the students who had 40 percent to 60 percent correct answers on the reading comprehension test, as compared to the students who had 80 percent to 100 percent correct answers on the same test. Although Level 4, 'reading fluently with full comprehension', is an absolute standard of performance agreed to by the MoE and language experts, fewer than 9 percent and 6 percent of the sample students belonged to Level 4 in 2010 and 2016, respectively. Therefore, the present study used the measure of proficient reader, which is equivalent to the *functional* reader and consists of students at Level 3 or above. In both EGRA 2010 and EGRA 2016, about 30 percent of students were classified as proficient readers.

Table 2.6. Reading Proficiency Level by Languages and Grades

Language	Region	Grade	Non-Reader	Reading slowly with limited comp	Proficient Reader: Reading with some fluency and comp	Proficient Reader: Reading fluently with full comp
			(Level 1)	(Level 2)	(Level 3)	(Level 4)
			Correct Words per Minute (CWPM) Measured by ORF			
Afan Oromo	**(Oromia)**	Grade 2	0	1-19	20-47	48
		Grade 3	0	1-29	30-57	58
Af-Somali	**(Somali)**	Grade 2	0	1-24	25-49	50
		Grade 3	0	1-24	25-54	55
Amharic	**(Amhara)**	Grade 2	0	1-29	30-49	50
		Grade 3	0	1-34	35-59	60
Sidamu-Afoo	**(SNNP)**	Grade 2	0	1-19	20-44	45
		Grade 3	0	1-24	25-52	53
Tigrinya	**(Tigray)**	Grade 2	0	1-20	20-54	55
		Grade 3	0	1-25	25-61	62

Source: EdDataII—Results of the Early Grade Reading Benchmarking Workshop in Ethiopia, RTI (2015)

3) Model controls

The present study included a set of control variables contained in both EGRA 2010 and EGRA 2016 to account for non-random sources of selection into preschool. The EGRA is a school-based assessment, thus background information was collected on students, teachers, and school principals by a contextual questionnaire. The questionnaire included a set of questions about student, parent, and family characteristics that potentially affected ECE exposure and students' learning outcomes; however, this provided relatively limited information as compared to household-based surveys, which included a report from caregivers. Hence, the control variables used for the present analysis were initially limited to the few measures available and to its comparability between EGRA 2010 and EGRA 2016, before being guided by theory and prior research on these factors.

Specifically, I first included child age and gender as demographic characteristic. Previous studies in LMICs and in Ethiopia have suggested that boys and older children are more likely to enrol in school and to present better academic performance (Lewin, 2009; Piper, 2010). Second, I included family characteristics that prior work has found to be associated with school participation and child outcomes in low-resource settings, including father's and mother's literacy, whether children speak the same language at home as they are taught in at school, and whether children have reading materials at home (Banerjee et al., 2008; Black et al., 2017; Seid, 2016). Notably, on the EGRA 2010, having reading materials at home was a stronger predictor of better learning outcomes than being wealthy (as measured by higher socioeconomic status [SES] index) (Piper, 2010).[42] Third, geographical covariates were included in the current analysis to consider the potential overlap with preschool access and resource availability (Rossiter et al., 2018; Sun, Rao, & Pearson, 2015; Woldehanna & Gebremedhin, 2012). These included schools located in an urban versus rural areas and dummy variables of five regions in Ethiopia.

As a limitation of the present study, the EGRA dataset has few measures for the SES or income levels of a family that prior research has found to be associated with school participation and child outcomes in LMICs (Black et al., 2017; Grantham-McGregor et al., 2007). EGRA 2010 collected data on a set of previously validated household asset indicators (e.g., ownership of a radio, television, phone, bike, car, or animal, or whether the house has electricity, a roof, and a floor), whereas EGRA 2016 stopped collecting any of these indicators because the administration agency raised the issue of the reliability of self-reported responses from young children in Grade 2 and Grade 3.[43]

Two variables collected in EGRA 2010 and EGRA 2016 were excluded from the current analysis: whether any family members help students with homework and whether a child was absent from school more than a week before the survey. These were excluded because they were regarded as post-treatment inputs to preschool attendance that occurred after a child entered primary school.[44] Moreover, a family's support for a child's homework has an issue of multicollinearity, a situation in which two or more explanatory variables are highly linearly related. For example, family support is highly correlated with other independent variables in the regression model, such as mother's literacy (0.30, $p < .01$) and father's literacy (0.34, $p < 0.01$). When multiple regressors are imperfectly multicollinear, the coefficients on these regressors will be imprecisely estimated, due to a large sampling variance (Stock & Watson, 2015).

[42] For example, having other books at home is related with 10.3 more correct words per minute (cwpm) in Oromia, much larger than 3.0 cwpm related to higher SES index group (Piper, 2010). Please note that SES index is only available for the 2010 cohort, not for the 2016 cohort, thus is not included in the present analysis.
[43] Based on the personal interview with EGRA 2016 administration agency.
[44] In some cases, the type of questions has been changed. In EGRA 2010, six questions were asked to students specifically about who provides support doing their homework, including father, mother, siblings, other relatives, or tutors. EGRA 2016 asked a single question on whether any household members help with their homework.

2.7 Empirical Strategy

2.7.1 Research Question 1: Probability-Probability Plot

My first research question offers alternative views on test score metrics that compare achievement gaps in a nonparametric framework, as proposed by Livingston (2006), Ho (2009), and Ho and Reardon (2012). Using ordinal 'proficiency categories', this approach focuses in particular on the comparison of test score *distribution* across different tests, rather than on the comparison of *mean* test scores. This approach is pertinent to the present analysis, which examined the link between changes in policies and changes in achievement gaps between preschool attendees and non-attendees over time.

The traditional achievement gap measures were concentrated on a difference between *group averages* or *standard deviation units* (i.e., effect sizes) and the *percentage-above-cut metrics* (i.e., differences in the share of students above the proficient level).[45] Ho and Reardon (2012), however, raised the theoretical and practical challenges of these conventional approaches which concern their 'transformation-dependence' and 'cut-score-dependence'.[46] First, average-based gap metrics, which rely on the assumption of equal-interval scale properties, are variable under plausible transformations of the test score scale (Ho & Reardon, 2012). If equal-interval differences are not supported at all levels of the test score distribution, which is often the case in educational measurement, 'nonlinear transformation becomes permissible and distortions of averages and Cohen-type effect sizes will result' (p. 5). When it comes to the estimation of gap *trends,* traditional approaches further assume that the equal-interval scale properties are maintained over time, which leads to bias in the estimates. Second, the percentage-based metrics rely on how to define cut score; this could be altered substantially under the different cut scores or confound the comparison of test score distributions, depending on the density of students adjacent to the cut score (see more in Ho, 2008; Holland, 2002).

[45] The importance of 'proficiency categories' was heightened under the No Child Left Behind policy in the U.S., as the federal government sets the cut-off scores for proficiency levels and requires the state to meet this categorical threshold.

[46] Given that some states in the U.S. only disclose results in terms of categorical achievement levels (censored data), the lack of standard distributional statistics (test score) is one of the challenges listed by Ho and Reardon (2012), but this is not the case with the present paper.

To overcome these shortcomings, alternative gap metrics were proposed, including graphs and statistics that 'share the property of invariance under monotone transformations of scales, thereby providing an attractive basis for comparisons across tests with different scales' (Ho, 2009, p. 202). Specifically, the probability-probability plot (PP plot hereafter) was used to present a transformation invariant comparison of a pair of test score distributions. The PP plots, which were first proposed by Wilk and Gnanadesikan (1968), were applied in the context of educational test gaps by Spencer (1983), Livingston (2006), Ho (2009), Ho and Reardon (2012). According to Ho (2009), this nonparametric representation of trends, gaps, and gap trends (TGGT) holds 'both theoretical and practical advantages for cross-test comparisons, particularly as they may help to encourage a *distribution-wide* perspective TGGT-based inferences' (p. 202). This provides a stronger basis for achievement gap comparisons over time and across different sub-geographical units (e.g., districts, provinces, or villages). The present study's novel approach to capturing *gap trends* during the policy change offers more comparable and comprehensive measures for the achievement gap between preschool attendees and non-attendees than any conventional approach.

2.7.2 Research Question 2(1): Ordinary Least Square and School Fixed Effects

1) Education production functions and ordinary least squares

To estimate the relationship between preschool education and student outcomes, the present study employed an education production function approach. The education production function, which draws from human capital theory (Becker, 1962), is concerned with measuring the productivity of the relationship between investments and a return in educational outcomes.[47] This approach, largely applied by economists, has focused on testing the hypothesis of a causal relationship between resource inputs into the education process—such as parental investment, school type, teacher quality, and school resources—and educational outcomes via human capital accumulated and improved productivity of the individual (for reviews, see Haveman & Wolfe, 1995; Hanushek, 2002). The education production function

[47] This framework is analogous with the Cobb-Douglas (1928) production function, which is the universal functional form of production growth that relates production inputs and output for the U.S. manufacturing sector in the early 1990s (Kleyn, Arashi, Bekker, & Millard, 2017). A production function identifies the maximum quantities of a particular good (or service) that can be produced per time period with various combinations of resources and with a given state of technology.

has been used as a framework to uncover child development, such as the determinants of children's cognitive achievement (Todd & Wolpin, 2003) and inequalities in educational achievement by socioeconomic background and race (Todd & Wolpin, 2007).

According to Todd and Wolpin (2003), the production function framework reflects theoretical notions that 'child development is a cumulative process depending on the history of family and school inputs and on innate ability' (p. 53). This approach offers a frame to map the interaction of family background and the current skill level of the individual child, along with investment at each age into the child's development and growth. The model requires the specification of a particular relationship between a set of inputs and the outputs. This is represented as follows:

$$Y_{it} = f(P_{it}, X_{it}, F_{it}, S_{it}, \epsilon_{it}) \qquad (1)$$

where Y_{it} represents the educational attainment for student *i* at time *t*, which is determined by a set of early childhood experiences (P_{it}), such as the quality of stimulation at home or any centre-based early learning from birth to before primary school entry; a set of individual characteristics (X_{it}), such as age and gender; a set of cumulative family (parent-chosen) inputs (F_{it}), such as parental education, wealth level, ethnicity, and home language; and school- or system-based inputs (S_{it}), such as school resources and management capacity. The key insight for the present study is that a child's learning productivity is partially determined by the parents' investment in sending the child to preschool in early childhood, which is part of P_{it}. Besides, although the individual characteristic X_{it} was often regarded as the individual's innate ability (Todd & Wolpin, 2003), more recent understanding of genetics (see Shonkoff & Phillips, 2000; Rutter, 2006) has revealed that the notion of a fixed genetic component can be misleading, given that an individual's genes interact with their early environment (Britton & Vignoles, 2017).

Using the Cobb-Douglass functional form expressed in equation (1), one can take the log of the production functions, which becomes a linear approximation in the parameters of interest. This approach is useful because it allows us to identify the degree of influence of *each* different input on child development. Following this framework, I used an ordinary least square (OLS hereafter) regression model to estimate the association between preschool attendance and students' academic achievement (test scores) that accounts for the various child- and household-level factors. The basis of my estimation strategy can be summarised as follows:

$$Y_i = \beta_0 + \beta_1 PRE_i + \beta_2 X_i + \beta_3 F_i + \epsilon_i \qquad (2)$$

where Y_i represents the academic achievement for student i measured by early grade reading assessment; PRE_i represents a binary variable of preschool attendance (a variable of interest in this study), an indicator that takes on a value of one if the child ever attended any form of pre-primary institution or zero if they did not. X_i and F_i represent covariates, each denoting student- and family-level characteristics, including regional dummies (i.e., region fixed effects that address between-region variations). β_1 to β_3 are the respective coefficients for these three vectors, and ε_i is an error term (residual) that captures unmeasured variables.[48]

In equation (2), the parameter of interest in this paper is β_1, which is supposed to capture the association between preschool attendance and students' academic achievement. To address my second research question—whether this association changed during the early learning reform in Ethiopia—the present analysis focused in particular on how this parameter β_1 changed between 2010 (t-1) and 2016 (t), denoting the pre- and post-reform periods. This can be expressed simply by the difference between β_{1t-1} and β_{1t}. Using two cross-sectional EGRA datasets with comparable information, I estimated equation (2) for the 2010 and 2016 cohorts separately in order to capture *trends* on whether the role of preschool in predicting early grade reading achievement had changed from before to after the reform.[49]

Selection bias of preschool attendance. My findings should be interpreted as the association between preschool attendance and students' academic outcomes. In the absence of experimental data randomly assigning children to preschool, I was unable to adequately control for non-random selection into preschool (e.g., parental motivation or local education policy)

[48] To account for the possibility of correlated errors across individuals nested in school, all models include robust standard errors clustered at the school level.

[49] Provided that Ethiopia's preschool expansion in 2011 created an exogenous source of variation in cohorts (i.e., a 2010 cohort who hadn't exposed to the reform versus a 2016 cohort who were exposed to the reform) and regions (i.e., a different pace of preschool expansion by region), I could have conducted a model using the difference-in-difference (DID) framework, which enables researchers to identify the causal inference. However, due to the data limitations, I could not meet the conditions of applying DID. The two EGRA datasets here are neither following the same set of children (panel) nor containing the same unit of analysis (e.g., district, village, or smaller administrative units like a woreda). Besides, there is a considerable regional difference in the preschool expansion and the phase of expansion is irregular between and within region, as it was heavily dependent on the capacity of regional/local governments and communities.

by the OLS model in equation (2).[50] Students who attended preschool tended to be different in both observable and unobservable ways from their peers who did not attend preschool. In Ethiopia, for instance, the key predictors of attending preschool were having a more educated caregiver, belonging to a household with greater wealth, living in an urban area, and being a first-born child (Vandemoortele, 2018), and no single factor explained all the determinants of preschool. To the extent that sources of selection into preschool were confounded by unobservable characteristics (ε_i) over and above those included in covariates (X_i and F_i), the concern for the unobserved variable bias remains; this is also known as selection bias or endogeneity. Therefore, coefficients presented in this study should be interpreted as indicating association, not as causal inferences.

2) School fixed effects versus random effects model

Another challenge to the straightforward OLS estimates is that children who attended preschool may be selected into different schools than other children. Because children with preschool experience are more likely to sort into a high-performing school than a low-performing school (Magnuson, Meyers, & Ruhm, 2004), a simple comparison to children without any preschool experience will be biased upward. This motivates the use of school fixed effects, which compares children who attended preschool with children in their schools who did not. Because the current analysis focused on the role of preschool in determining student outcomes no matter what subsequent school experience they had, the model including a school fixed effect could ensure that important school-level variance (e.g., school resources) was adjusted for the estimation.

The estimation of school fixed effects could be best understood in the multi-level (or mixed) model. In principle, any analysis including student achievement needs to reflect the hierarchical nature of the data, where students are nested within schools. It is well established that students' achievement in the same school is likely to be clustered, due to the influence of unmeasured school characteristics such as schoolwide policies and leadership (see Goldstein, 1995; Raudenbush & Bryk, 2002; Snijders & Bosker, 1999). Using multi-level models accounts for students' clustering in schools while allowing student- and school-level characteristics to be

[50] The OLS model holds the assumption that all factors in the unobserved error term (ε) are not correlated with the explanatory variables, expressed as E (ε | *PRESCHOOL*) = 0.

included together (Clarke, Crawford, Steele, & Vignoles, 2015). Because multi-level models enable researchers to assess between- and within-school variance in student outcomes, estimation of multi-level regression models can be done by treating school effects as either *random* or *fixed*. In the following section, I review each approach to identify the best model for the current study, based on their strengths and weaknesses.

The following equation (3) represents a *random* intercept multi-level model that reflects a two-level nested structure:

$$Y_{is} = \beta_0 + \beta_1 PRE_{is} + \beta_2 X_{is} + \beta_3 S_s + u_s + \varepsilon_{is} \qquad (3)$$

where Y_{is} represents academic achievement of student *i* nested in school *s;* PRE_{is} denotes preschool attendance; X_{is} is covariates representing student- and family-level characteristics; and S_s represents school-level characteristics (that vary between schools). Unlike the OLS model (eq. (2)), the multi-level model contains a composite residual that sums two distinct error terms: u_s is the school-level residuals and ε_{is} is the student-level residual.[51] This also can be expressed as:

$$Y_{is} = (\beta_0 + u_s) + \beta_1 PRE_{is} + \beta_2 X_{is} + \beta_3 S_s + \varepsilon_{is} \qquad (4)$$

By a simple re-ordering of the terms in the model itself, equation (4) can explain why this model is referred to as a *random* intercept multi-level model. Essentially, the inclusion of school-level residuals u_s in the model leads us to provide each school with its own 'random' intercept, represented by $(\beta_0 + u_s)$. When we fit this random effect model to data, we do not estimate each of the school-specific intercepts but the school-level residual variance, which is a summary of the variability in the *school-mean value of the outcome* from school to school (Murnane & Willett, 2010). This is often referred to as between-school variance, which summarizes the scatter in the outcome among schools. However, in the random-effects models, the probability of individuals attending preschool depends only on observed school characteristic S_s (e.g., whether the school has an O-Class or not). In other words, the random

[51] The school-level residual u_s is the sum of (a) all effects on academic performance of the school-level factors (which are *correlated* with preschool attendance) in F^{school}, and (b) the effects of all other school-level influences on academic performance which are *uncorrelated* with the preschool attendance. Thus, the random effects assumption on 'u_s should not be correlated with $Preschool_{is}$' can be failed if F^{school} is not empty.

effect assumption holds only when unobserved school effects are not correlated with the probability of attending preschool, expressed as (E $(u_s | \, PRE_{is}) = 0$).

Nevertheless, in many real settings, the probability of attending preschool also depends on unobserved school characteristics. For instance, school principals could try to enrol as many children in O-Class as possible if higher enrolment attracts additional financial resources. When the random effects assumption is unlikely to hold, the school fixed effect model offers a potential solution to account for unobserved school effects by containing an intercept parameter for each school. In fitting the school fixed effect model to the data from the EGRA 2016, there would be 225 such intercepts, one for each of the 225 schools in the sample. Equation (5) exemplifies the theory behind the fixed effect assumption:

$$Y_{is} = (\beta_0 + \beta_2 S_{2s} + \beta_3 S_{3s} + \dots + \beta_{225} S_{225s}) + \beta_1 PRE_{is} + \beta_2 X_{is} + \beta_3 S_s + \varepsilon_{is} \quad (5)$$

The equation creates a set of dichotomous variables, S_1 through S_{225}, to represent the 225 schools in the sample, setting each of these dummies equal to 1 when the child belongs to that school, to 0 otherwise (as the reference category, the first school S_1 has dropped in the model). After the school-level variance u_s has been removed from the outcome variability, the only residual remaining in the model is the child-level residual variance ε_{is}. It often is referred to as within-school variance, which describes the scatter in the outcome from student to student within each school. Because this child-level residual satisfies OLS assumptions implicitly, we do not need any further assumption to fit a fixed effect model (Clarke et al., 2015; Murnane & Willett, 2010).[52] The identifying assumption is that selection into preschool among students of the same school is uncorrelated with the unobserved school-level determinants of outcomes, which are stable elements over time.[53]

The choice of the appropriate model—random versus fixed—is often driven by discipline tradition; however, the decision should be guided by the research context and data-specific characteristics (Clarke et al., 2015). In my empirical questions, the primary interest is students'

[52] Unlike the random-effects assumption, it does not matter if the fixed effects that present the grouping are correlated with other predictors in the model at any level, because regression analysis is designed to permit predictors to be correlated (Murnane & Willet, p. 133).
[53] School fixed effects only account for time-invariant unobserved differences, thus the estimate can be biased if any time-variant difference occurred to the school or community (e.g., shocks by drought, flood, or political clashes).

characteristics (i.e., preschool attendance) rather than the characteristics of the schools they attended at the primary level. In this regard, the school fixed effect model is appropriate, as it allows me to control for unobserved, time-invariant between-school differences which may sort preschool attendees and non-attendees during the transition to primary school. This approach is useful for the current study where the data were quite limited to adjust for the effects of preschool assignment from multiple sources.

One disadvantage of the fixed effects approach as compared to the random effect approach is 'efficiency'. Given that the number of regression parameters will increase vastly by the inclusion of a dichotomous predictor of each school, the degrees of freedom are reduced, then the standard errors tend to increase.[54] Moreover, given that schools are used as their own controls, a school fixed effects model requires certain within-school variability in the variable of interest (preschool attendance). If there is little variability within schools, the standard errors from fixed effects models may be too large to tolerate; thus, this will be discarded from the estimation (Allison, 2009).

Importantly, it should be acknowledged that neither a school fixed effect nor random effect model will address the unobserved characteristics of students and parents. All the estimates using the multi-level model thus should be recognised as an association, although the fixed effect model can provide better estimates than OLS or a random effect model in the present study. In fact, prior studies on the effect of preschool attendance used the 'household' or 'sibling' fixed effect model to account for the unobserved characteristics of parents and household (Berlinski et al., 2008; Bietenbeck et al., 2017; Currie & Thomas, 1995; Deming, 2009; Garces, Thomas, & Currie, 2002). This approach could be more robust than the school fixed effect model; however, EGRA is the school-based assessment and such information at the household level is not available for the current study.

[54] The standard error is a measure of how representative of the population a sample is likely to be (the accuracy with which a sample represents a population). Therefore, having a large standard error relative to the mean indicates that there is a lot of variability between the means of different samples, and therefore the sample might not be representative of the population.

2.7.3 Research Question 2(2): Logistic Regression, Predicted Probabilities, and Marginal Effects

The next research question continues to address the relation between preschool attendance and early grade reading outcomes by using different outcome variables on the probability of being a non-reader or a proficient reader. These outcome variables take only two possible values, that is, either non-reader or not, either proficient reader or not; hence, I used a *multivariate* logistic regression model for this analysis. Initially, given that reading proficiency was categorised into four levels, the *ordered* logistic or *multi-nominal* logistic models were considered but dropped, due to concerns about the ordinality of the dependent variable and interpretability. First, the assumption of an *ordered* logistic model is that the interval among the categorical dependent variables (more than two) should be of equal distances on a single, underlying dimension (Long & Freese, 2014). However, this is not the case for the reading proficiency categories in Ethiopia, as shown in Table 2.6. Five mother tongues have their own distinctive metrics, which contain varied benchmark ranges for each proficiency category and by grade. For example, on reading proficiency levels in Amharic, the benchmark ranges for Levels 2 and 3 are 1-29 and 30-49 for Grade 2 and 1-34 and 35-59 for Grade 3, respectively.

Alternatively, a *multi-nominal* logistic model can be considered where the categories are assumed to be unordered. However, the biggest challenges of a *multi-nominal* logistic model are that the model includes many parameters and so it is easy to be overwhelmed by the complexity of the results (Long & Freese, 2014). This complexity is compounded by the nonlinearity of the model, which adds more complications in interpretation. Provided that the *multi-nominal* logistic model essentially fit separate binary logistics for each pair of outcome categories, I decided to use the most meaningful pair of outcome categories in the EGRA context (i.e., non-reader: Level 1 vs. Levels 2, 3, 4; proficient reader: Levels 1, 2 vs. Levels 3, 4) using the *multivariate* logistic regression model.[55] Lastly, one can consider the benchmark for the reading comprehension test (80%, or 4 out of 5 correct answers, regardless of language) to avoid high dependency on ORF scores. However, student performance on the reading comprehension test was generally too low, as only 6 percent of Grade 2 and Grade 3 students

[55] The proportion of students at Level 4 was less than 9 percent and 6 percent in 2010 and 2016, respectively, thus division by Levels 1, 2, 3 vs. Level 4 was excluded.

reached this benchmark. About 78 percent and 84 percent of students could reach even a 50 percent reading comprehension level in 2010 and 2016, respectively.

When using nonlinear models, the simple interpretations of the estimated parameters that are applied to linear models are no longer valid. In nonlinear models, the effect of a change in a variable depends on the values of all variables in the model. For instance, the curves are not parallel in the probability density function used in the nonlinear model, thus the magnitude of the difference in the predicted probability y at a binary independent variable $d = 1$ compared with $d = 0$ depends on the values of x (a continuous independent variable), where the difference is computed (Long & Freese, 2014, p. 136). Because of this nonlinearity, no single method of interpretation can fully describe the relationships among independent variables and dependent variables. Instead, a series of post-estimation explorations, largely based on predictions, are needed to uncover the most important aspects of these relationships. By computing predicted or expected values for hypothetical or prototypical cases, researchers can covey more tangible and practical significance of the findings (Cameron & Trivedi, 2010; Long & Freese, 2014; Williams, 2012).

Therefore, I documented the results of the logistic regression in terms of *odds ratio, marginal effects,* and *predicted probabilities.*[56] *Odds ratio* is the multiplicative change in probability for a unit increase in explanatory variable, holding other variables constant. For the binary measures such as non-reader, odds ratio represents the relative change in the probability of being a non-reader when preschool attendance is measured against a reference category—0 (no preschool) and 1 (preschool). Generally, odds ratio > 1 indicates increased occurrence of outcome variables and odds ratio < 1 indicates decreased occurrence of outcome variables. It is helpful to look at how much the odds ratio deviates from 1 (no association), for example, an odds ratio of 0.75 (less than 1) means that the outcome is 25 percent less likely to happen in one group.[57] When interpreting the outcome variables in the present study, odds ratio < 1 is positive results for ‘reducing’ the probability of being a non-reader, whereas odds ratio > 1 is positive results for ‘increasing’ the probability of being a proficient reader.

[56] Given that the standard outputs of regression coefficient in nonlinear models do not provide a unique interpretation of the relationship of the explanatory variables with the outcomes, I didn’t present the regression coefficients in this paper (the result is available upon request).
[57] Likewise, an odds ratio of 1.33 (greater than 1) means that the outcome is 33 percent more likely to occur in one group.

Nevertheless, the interpretation of the odds ratio is not straightforward. The odds ratio indicates the direction of the relationship and its statistical significance; however, this has little value in conveying the magnitude of effects (Long & Freese, 2014). Moreover, the odds ratio could limit the reflection of nonlinearity in the model, as it assumes to hold other variables constant regardless of their levels, and a constant factor change in the odds does not imply a constant change in the probability. To facilitate an interpretation of the results in more practical terms, I presented marginal effects and predicted probabilities.

Marginal effect (differences in predicted probabilities) is the additive change in predicted probability for a unit increase in explanatory variable, holding other variables at specific values (Long & Freese, 2014). The term 'effect', however, does not necessarily mean causal inference. The 'marginal effect' or simply 'effect' of a variable 'preschool (X)' on 'learning outcomes (Y)' is, in fact, the marginal variation (partial derivative) of Y associated with a unit change in preschool attendance X from 0 to 1.

Unlike the marginal effects in the linear regression that equal the relevant slope coefficients, the calculation of the marginal effects in the nonlinear regression remarkably varies according to the value at which the other variables are set (Cameron & Trivedi, 2010). This can be calculated as 'average marginal effects' or 'marginal effects at the means', which are widely used approaches across disciplines. In the present study, average marginal effects were used, given that theory does not support the selection of specific fixed values used in marginal effects at the means. Average marginal effects (AMEs), or average predicted probabilities, are the expected probability of a person with average characteristics. In other words, the average predicted probability of being a non-reader for those who attended preschool was computed for each child, using that child's actual observed values of all the other explanatory variables when preschool was set as equal to attending preschool, and the predicted values from each case were then averaged.

By comparison, marginal effects at the means (MEMs), or predicted probabilities at the means, are the expected probability of a person holding all other predictors fixed at their mean values when preschool was set as equal to attending preschool. In the current analysis, for example, the average individual at their mean values indicated that the student was 9.84 years old, 50 percent were female, 95 percent used the same language at home, 44 percent had reading materials at home, 72 percent had a father who can read and write, and 47 percent had a mother who can read and write, which is an implausible individual. A common criticism of the MEMs is that there typically is no actual case in the dataset for which all variables equal the mean.

Taking into account this limitation of MEMs, some researchers (e.g., Bartus, 2005; Cameron & Trivedi, 2010) prefer to use AMEs, which could be the best summary of the effect that averages the individual effects across all cases in the sample. Nevertheless, it should be noted that researchers are divided as to which model produces superior estimates. No matter how 'average' is defined, two common approaches produce only a single estimate of the marginal effects, and this can obscure differences in effects across cases (Williams, 2012). Bear this caveat in mind: when interpreting results in the current analysis, it is important to remember that these are AMEs and average predicted probabilities that could be interpreted as the average size of the effect in the sample.

2.8 Results

2.8.1 Descriptive Statistics: Sample Characteristics and ECE Participation

1) Sample characteristics

Table 2.7 presents the sample descriptive statistics, including the overall sample characteristics and whether the child had ever attended preschool. On average, children were ten years old at the time of the assessment. Among the sample evenly stratified by gender and grades, 48 percent and 51 percent of female students reported that they attended preschool in 2010 and 2016, respectively.[58] On average, about 80 percent of children were living in rural areas; among students who attended preschool, the percentage living in rural areas increased from 66 percent in 2010 to 71 percent in 2016, while 85 percent of students from both cohorts who didn't attend preschool were living in rural areas.[59] In total, students who attended preschool were more likely to have a mother tongue textbook, have reading materials at home, and have a father or mother who could read and write than those who didn't attend preschool, while overall gaps between the two groups widened from 2010 to 2016. For instance, the differences in mother's

[58] These corresponded to the MOE's education annual statistics reporting no gender disparity in pre-primary.
[59] The urban-rural composition in Table 2.7 is computed by (preschooler in rural)/(total rural population), whereas Table 2.9 is computed by (preschooler in rural)/(total preschooler in urban and rural).

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.[60] Although children who didn't attend preschool were more likely to use the same language at home as that used for instruction than their peers who attended preschool, these differences were not statistically significant. Overall, the descriptive picture suggests that families that chose to send their child to preschool in 2010 were not necessarily similar to the ones who did so in 2016. The substantial change in the composition of attendees versus non-attendees during the massive expansion of preschool, as well as differences between the two groups, highlights the need to account for baseline differences when estimating the role of preschool attendance in predicting child academic outcomes.

Table 2.7 also reports the average difference in early grade reading achievement based on preschool attendance between 2010 and 2016. Across eight outcome variables measured by the EGRA sub-tasks and the proportion of non-readers and proficient readers, students who attended preschool showed better performance on average than students who did not (except the proportion of non-readers in 2010). Notably, these achievement gaps widened between 2010 and 2016—that is, preschool attendees read 5.8 more correct letter sounds per minute (clpm) than non-attendees in 2010, and this gap had grown to 10.5 clpm between preschool attendees and non-attendees in 2016. It should be noted that, considering the fundamental differences among the five languages, the direct comparison of average test scores is not a preferred approach using EGRA (AIR, 2016; Piper, 2010).[61] The descriptive summary on *mean* difference provides rough estimations of trends in early grade reading achievement in 2010 and 2016; the in-depth analysis by regions/languages is presented in the next section.

[60] Note that there is an overall improvement of adult literacy rates in Ethiopia over this period. According to the Ethiopia's Demographic and Household Survey (DHS, 2011, 2016), between 2011 and 2016, females (among women aged 15-49) who can read a whole sentence or part of a sentence increased from 27.2 percent to 39.6 percent, and males (among men aged 15-49) who can read a whole sentence or part of a sentence increased from 48.6 percent to 62 percent.
[61] The descriptive statistics by regions/languages can be found in the EGRA country report by RTI (2010) and AIR (2016) and is also available upon request.

Table 2.7. Descriptive Statistics of Control Variables

						2010			2016		
	2010 average		2016 average		*Dif*	Pre (a)	No-Pre (b)	(a)-(b)	Pre (a)	No-Pre (b)	(a)-(b)
	m	(SD)	*m*	(SD)		*m*	*m*	*Diff.*	*m*	*m*	*Diff.*
Variable of Interest											
Preschool Attendance	0.14	(0.35)	0.38	(0.49)	***	-	-	-	-	-	-
Covariates											
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**
Rural	0.82	(0.38)	0.79	(0.41)	-	0.66	0.85	-0.19**	0.71	0.84	-0.13***
MT textbook	0.75	(0.43)	0.71	(0.45)	-	0.80	0.74	0.06	0.77	0.68	0.09***
Book at home	0.21	(0.41)	0.44	(0.50)	***	0.30	0.20	0.10***	0.53	0.38	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***
Father's literacy	0.52	(0.50)	0.72	(0.45)	***	0.64	0.50	0.14***	0.81	0.66	0.15***
Same language of instruction	0.90	(0.30)	0.94	(0.23)	-	0.85	0.91	-0.06	0.94	0.95	-0.01
Outcome Variables											
ORF (cwpm)	21.78	(21.35)	21.25	(20.93)	-	23.79	21.42	2.37	23.85	19.66	4.19***
Letter sounds recognition (clpm)	45.35	(31.29)	47.61	(31.28)	-	50.24	44.48	5.76*	54.11	43.64	10.47***
Familiar words recognition (cwpm)	20.85	(19.43)	22.78	(21.31)	-	22.58	20.49	2.09	26.14	20.73	5.41***
Invented words recognition (cwpm)	13.72	(13.68)	15.57	(15.29)	-	15.61	13.44	2.17**	18.09	14.04	4.05***
Reading comprehension (% of correct)	24.00	(27.00)	22.59	(27.70)	-	28.02	23.29	4.73*	25.57	20.76	4.81***
Listening comprehension (% of correct)	55.82	(29.95)	74.31	(26.09)	***	62.28	54.47	7.81**	77.15	72.57	4.58***
% of non-reader	0.32	(0.47)	0.26	(0.44)	**	0.34	0.32	0.02	0.19	0.30	-0.11***
% 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: (1) All figures were weighted by sample weight, given that the weighted descriptive values could accurately measure the true value in the population (Solon, Haider, & Wooldridge, 2013); (2) 'Dif' column shows the results of test for difference in proportions or means with different samples (EGRA 2010 and EGRA 2016). *** p<0.01, ** p<0.05, *p<0.1
Source: EGRA Dataset 2010, 2016, USAID

2) ECE participation: A shift from an elite to a mass system

With respect to preschool access, approximately 14.2 percent and 37.9 percent of children were reported to have attended any form of ECE, nearly tripling between 2010 and 2016 (Table 2.7). Table 2.8 provides further details on the expansion of pre-primary education by region during the reform period. As shown in Table 2.8, there were significant regional variations in the growth of preschool coverage. The most notable expansion was observed in SNNP (Sidamu), where preschool participation soared from 18.2 percent to 64.4 percent over the six-year period. In contrast, there was no change in preschool participation in Oromia between 2010 and 2016. Table 2.9 illustrates these trends, divided by urban and rural residence. In 2010, the proportion of children who attended preschool was, on average, 27.4 percent for those living in an urban area and 11.4 percent for those living in a rural area. In 2016, this proportion increased significantly in both locations: 52 percent of children from urban and 34 percent of children from rural areas attended preschool before they entered formal schooling. Noticeably, preschool attendance in rural areas showed a steeper increase between 2010 and 2016 than in urban areas, but it still lagged behind preschool coverage in urban areas.

Table 2.8. Proportion of Children Who Attended Preschool (EGRA)

Region	EGRA 2010				EGRA 2016			
	Pre	No-Pre	Total	**% (weighted)**	Pre	No-Pre	Total	**%(weighted)**
Tigray	186	1,351	1,537	**12.1% (12.6)**	769	940	1,709	**45.0% (50.0)**
Amhara	158	2,101	2,259	**7.0% (6.1)**	402	1,346	1,748	**23.0% (22.0)**
Oromia	491	1,951	2,442	**20.1% (21.9)**	366	1,383	1,749	**20.9% (23.6)**
Somali	127	1,036	1,163	**11.0% (10.0)**	284	1,068	1,352	**21.0% (27.2)**
SNNP (Sidamu)	283	1,437	1,720	**16.5% (18.2)**	1,168	606	1,774	**65.8% (64.4)**
Total	1,245	7,876	9,121	**13.6% (14.2)**	2,989	5,343	8,332	**35.9% (37.9)**

Note: (1) In the parenthesis, weighted values are presented.
Source: EGRA Dataset 2010, 2016, USAID

Table 2.9. Proportion of Preschool Attendance by Urban-Rural (EGRA)

Region	EGRA 2010				EGRA 2016			
	Total Urban Sample (N)	Pupil attended Preschool (%)	Total Rural Sample (N)	Pupil attended Preschool (%)	Total Urban Sample (N)	Pupil attended Preschool (%)	Total Rural Sample (N)	Pupil attended Preschool (%)
Tigray	197	31.4	1,340	11.1	223	75.7	1,486	42.8
Amhara	238	13.8	2,021	5.7	142	42.0	1,606	18.8
Oromia	540	39.3	1,902	12.6	190	35.3	1,559	21.1
Somali	515	10.5	648	9.7	193	19.2	1,159	29.1
SNNP (Sidamu)	197	9.4	1,523	19.0	314	64.9	1,460	64.1
Total	1,687	27.4	7,434	11.4	1,062	52.0	7,270	34.1

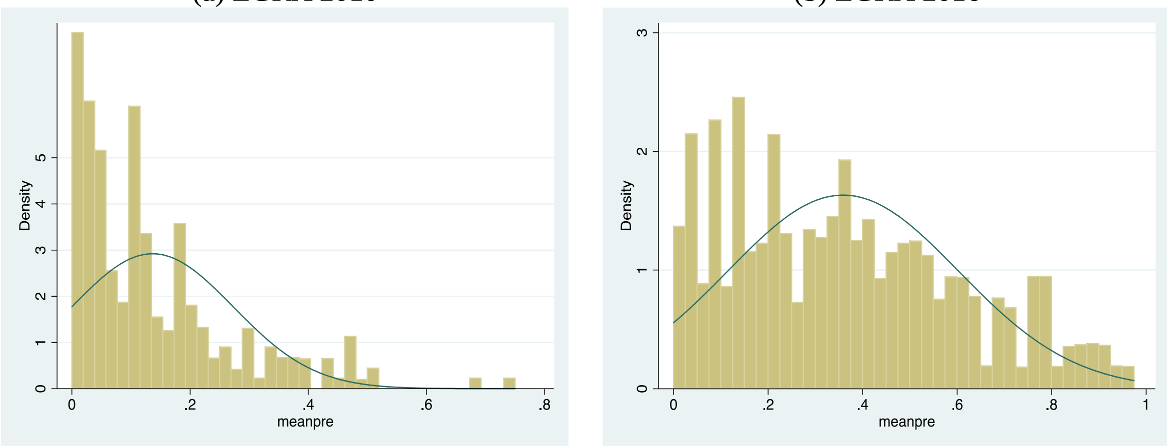
Note: (1) In EGRA 2010, Somali used convenience sampling (similar portion of urban and rural sample) due to security issues; (2) In EGRA 2016, random sampling was conducted after excluding the areas listed as priority zones by the Emergency Education Cluster report (UNICEF, 2016) across five regions (AIR, 2016).
Source: EGRA Dataset 2010, 2016, USAID

Figure 2.7 displays the distribution of the school-level average of pupils enrolled in any form of preschool in both cohorts. This figure depicts how access to pre-primary education in

Ethiopia shifted from the *elite* system to the *mass* system during the reform. In 2010, more than half of schools (125 out of 237) had less than 10 percent of students who had ever attended preschool, which shows a skewed right distribution in preschool enrolment, with an average of 14 percent.[62] Conversely, in 2016, after the massive expansion of O-Class, the distribution of preschool enrolment shifted close to a normal distribution. The average preschool enrolment almost tripled to 38 percent in 2016. About 30 percent of schools (62 out of 225) reported that more than half of students entered primary school after having attended preschool, while about one-fifth of schools still had less than 10 percent of students who ever attended preschool.

Figure 2.7. Kernel Density of School-Level Average of Pupils Who Enrolled in Preschool

(a) EGRA 2010 (b) EGRA 2016



Note: The figures include five regions/languages in the sample of the present study.
Source: EGRA Dataset 2010, 2016, USAID

In addition to preschool participation at the school level, I further scrutinised whether these sample schools from the EGRA 2016 actually had O-Class by matching information from EMIS.[63] Surprisingly, among the 225 sample schools, 147 primary schools (or 65%) already had O-Class attached to the school, which far exceeded the share of students who had ever attended preschool (38%). This was reaffirmed by the proportion of primary schools having O-Class within the same woreda. This proportion within the same community reached an average of 70 percent, which implies that the availability of O-Class was much higher than actual participation. Consistent with national patterns, this implies the limited capacity of the existing O-Class, which had only one or two classrooms attached to the primary school. As far

[62] A skewed right distribution indicates that the right tail (higher values) is much longer than the left tail (small values).
[63] The availability of O-Class is from EMIS 2014-2015 and 2015-2016, when the EGRA 2016 sample schools were selected.

as school principals in the Tigray region were concerned (personal interviews), the existing O-Class supply has not been sufficient to cater to *all* 6-year-old children in the community; they could serve only one-third, or less than half, of preschool-eligible children. The dissonance between demand for and supply of O-Class could have multiple causes (e.g., multi-aged O-Class, lack of awareness), which could negatively affect the equitable access to quality pre-primary education.

3) Determinants of preschool attendance

Before getting into the main research questions, it is important to identify the factors relating to children's preschool attendance, which may have changed during the massive expansion of preschool. Understanding the characteristics of preschool attendees not only sheds light methodologically on the potential direction of omitted bias, it also has policy implications as it reveals unequal access to preschools across different sociodemographic groups. Table 2.10 presents the results of a logistic regression analysis that was undertaken for each EGRA cohort. It presents the odds ratio, which indicates the direction and statistical significance of the relationship, and average marginal probabilities to ease interpretation. For example, the average marginal probability of attending preschool for females was calculated for each child, using that child's values for all the other explanatory variables, when sex is set equal to female.

The results in Table 2.10 show that family characteristics played a significant role in determining preschool attendance. First, the child's age and gender were not predictive of preschool attendance, suggesting little gender bias in preschool enrolment at an early age. Second, father's and mother's literacy and having reading materials at home were strongly related to a higher probability of attending a preschool, while these marginal probabilities nearly doubled between 2010 and 2016. Among others, the variable on having reading materials at home was the strongest predictor for preschool attendance: this was associated with a 6 and 10 percentage point higher probability of attending preschool in 2010 and 2016, respectively, compared to peers who did not have any reading materials at home. With regard to parental literacy, children with a literate mother had a 3 and 10 percentage point higher probability of attending preschool than children with an illiterate mother in 2010 and 2016, respectively. Mother's literacy was a stronger determinant of preschool attendance than father's literacy. Third, as expected, urban and rural residency was a strong predictor for preschool attendance. In both 2010 and 2016, children living in rural areas were about 9 percentage points less likely to attend preschool than those living in urban areas. Two other variables—the same language being used at home and at school and having mother tongue textbooks—were found to be non-significant predictors of preschool attendance in 2010, yet they became significant predictors in 2016: using the same language at home and at school was related to a lower probability of preschool attendance, and having mother tongue textbooks was related to a higher probability of preschool attendance in 2016.[64] These results further support the descriptive patterns observed in Table 2.7, that children from families with relatively advantaged backgrounds were more likely to attend preschool in Ethiopia.

Table 2.10. Determinants of Attendance in Preschool

	EGRA 2010			EGRA 2016		
	Odds Ratio	(SE)	Marginal Prob.	Odds Ratio	(SE)	Marginal Prob.
Age	0.90	(0.13)	-0.01	0.76	(0.16)	-0.05
Female	0.87	(0.09)	-0.02	1.04	(0.07)	0.01
Father's literacy	1.24**	(0.16)	0.02	1.33***	(0.11)	0.05
Mother's literacy	1.36***	(0.17)	0.03	1.71***	(0.12)	0.10
Reading materials at home	1.72***	(0.22)	0.06	1.74***	(0.12)	0.10
Rural location	0.45***	(0.06)	-0.09	0.65***	(0.06)	-0.08
Same language home/school	0.94	(0.11)	-0.01	0.69***	(0.09)	-0.08
Have MT textbook	0.99	(0.14)	-0.01	1.36***	(0.11)	0.06
Constant	1.04	(0.91)		27.60	(31.0)	
Grade dummies	Yes			Yes		
Region dummies	Yes			Yes		
Observations	9,121			8,332		

Note: (1) All models include sampling weight; (2) Robust standard errors in parentheses; (3) Marginal probabilities present average marginal effects (AME). *** p<0.01, ** p<0.05, *p<0.1
Source: EGRA Dataset 2010, 2016, USAID

2.8.2 Research Question 1: A Shift in Test Score Distribution during the Reform

Using the methods proposed by Livingston (2006), Ho (2009), and Ho and Reardon (2012), I addressed the following research question: *What is the difference in the test score distribution of second- and third-grade students' reading achievement, as measured by oral reading fluency, between preschool attendees and non-attendees before and after the early learning reform in Ethiopia?* Figure 2.8 presents the gap trends (metrics) with transformation invariant

[64] Although it is hard to find a clear explanation for the negative relationship between preschool attendance and same language use at home and school, 'using different language at home and school' seems to be a more urban phenomena (e.g., using official language Amharic at school and use mother tongue at home) from the EGRA sample. For example, 13 percent of children living in urban areas reported that they use different languages at home and school, while this occurs for 7 percent of children living rural areas.

properties through the *probability-probability* plots. The PP plots, which are derived from a pair of test scores on the same score scale, depict the differences not only in the mean scores or in selected percentiles, but anywhere in the score distributions over the full range of values of each variable. The PP plot allows researchers to determine the extent to which the comparison between the two groups differs from one test to the other in a comparable manner. This plot can be used even if the variables are measured in completely different and noncomparable units (Livingston, 2006).

As shown in Figure 2.8, the PP plot is best described by considering the two cumulative distribution function (CDF) figures. Considering the ordinality and arbitrary normalization of test scores, the reliable comparison of two groups in the sample can be obtained only if looking at the CDFs of their performance (Spencer, 1983; Bond & Lang, 2013). In CDFs presented on the left-hand panel of each cohort, $F_{pre}(x)$ and $F_{no-pre}(x)$ on the vertical axis (Y) denote the proportions of students at or below a given score *x* on the horizontal axis (X) in the *preschool* and *non-preschool* groups, respectively. The test score distributions in the CDF are generally labeled as a higher-scoring reference distribution, F_{pre} (blue line), and a lower-scoring focal distribution, F_{no-pre} (red line).[65] To illustrate, in the 2016 EGRA cohort, for the average cut score of reading proficiency Level 3 at 28 cwpm, 64.2 percent of the *preschool* group was at or below Level 3, whereas 72.1 percent of the *non-preschool* group was at or below Level 3.[66]

The right-hand panel of each cohort is the corresponding PP plot that shows the proportion of the *non-preschool* group below given percentiles of the *preschool* group, expressed in a ROCFIT plot.[67] This PP plot is generated by obtaining all paired cumulative proportions across the score scale underlying the CDFs (left-hand panel). For instance, the paired cumulative proportions at the reference point above (0.64, 0.72), which are derived from reading

[65] The CDF curves are not smooth, but the direction and amount of the differences between them is fairly clear.
[66] There is no single cut score for Level 3 or Level 4 on the Ethiopia EGRA, due to different benchmark settings by languages and by grade. Provided that the cut score for Level 3 varied from 20 to 35 correct words per minute, the average value of 28 was used for the exemplar reference point above.
[67] ROCFIT fits maximum-likelihood receiver operating characteristic (ROC) models assuming a binormal distribution of the latent variable (Pepe, 2003, citing from Stata Manual). This plot can be generated by Stata command ROCFIT. To facilitate plotting, each of the variables should be a discrete measure of a set of ordered categories. Ho and Reardon (2012) recommends ROCFIT as the best approach out of six candidates for the estimation of gaps under the censored data scenarios. Originally, ROCFIT curve was motivated by signal detection theory to medicine and psychology and allowed for the visualization of the trade-offs of Type I and Type II errors across a continuum of decision thresholds. The Lorenz Curve (Lorenz, 1905) is a case of a PP plot from economics, where the vertical and horizontal axes are the cumulative proportions of income and households ranked by income, respectively.

proficiency Level 3, can be found in the right-panel plot. 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. No matter how the scale is stretched or transformed horizontally, the intersections of the CDFs with this vertical slice will keep the same values' (p. 213). In other words, apart from using any scale information, the construction of the PP plot is solely based on paired cumulative proportions; thus, all statistics generated from the PP plot are invariant to any transformation of scale. Given that the full CDFs are known for both groups in the present study, the PP plot can generate nonparametric gaps statistics straightforwardly (Ho & Reardon, 2012).

In terms of the interpretation of the PP plot (right-hand panel of each cohort), which is formed by connecting a series of data points, the horizon axis (X) of the data point indicates the percentage of the preschoolers who attained that score or below; and the vertical axis (Y) of the data point indicates the percentage of the non-preschoolers who attained that same score or below. If the percentages from the two groups were equal, the data point would lie on the diagonal line (Y=X). The greater the difference between the two groups, the farther the data point departs from the diagonal line, as expressed by the larger bulge (Livingston, 2006). The area under the curves from the PP plot can be interpreted as the probability that a randomly drawn preschooler scored higher than a randomly drawn non-preschooler, while a probability of 0.5 represents no trend or gap. The area under the curves is considered an effective measure of the inherent validity of a diagnostic test.

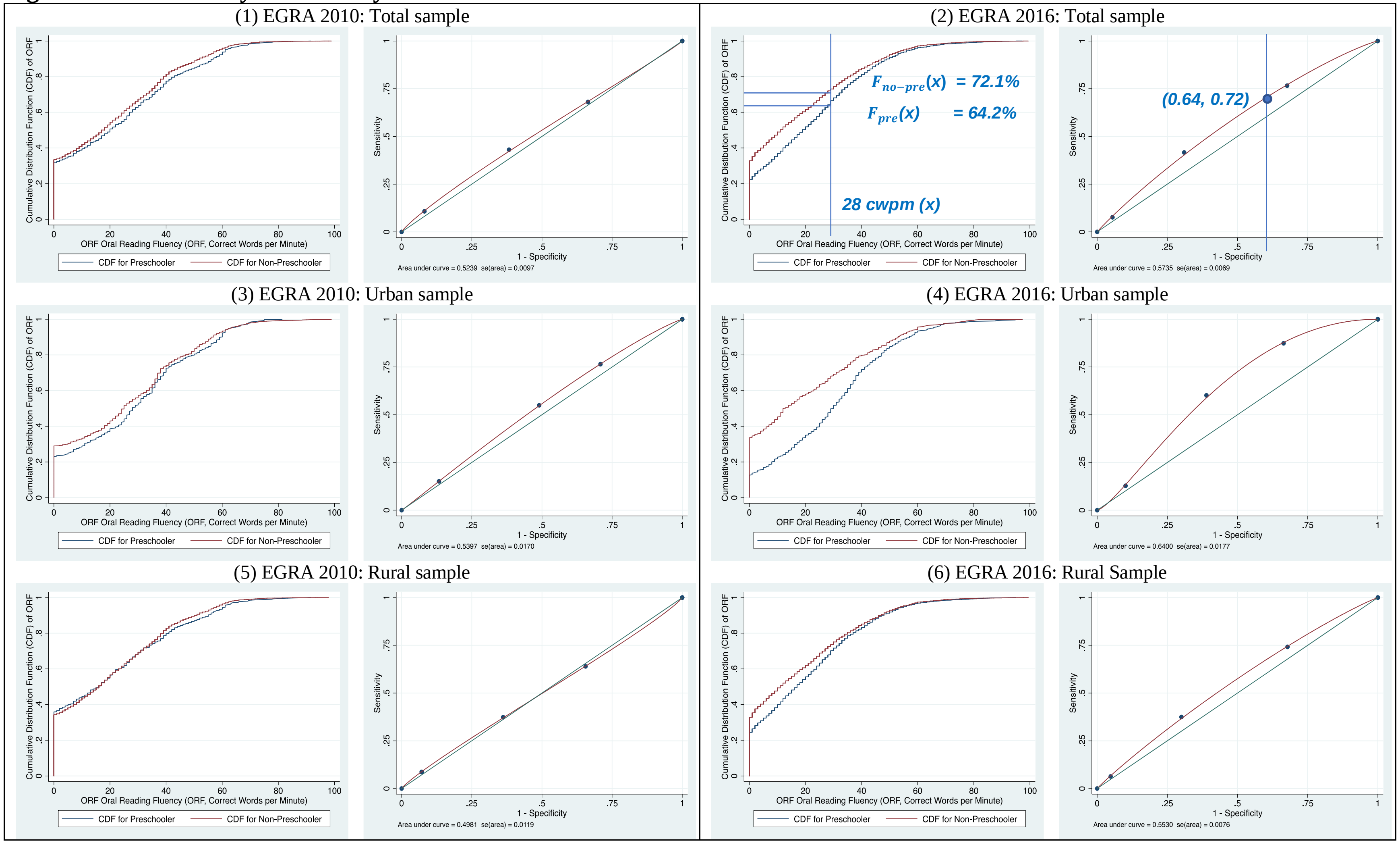
Specifically, the PP plots in Figure 2.8 capture the achievement gaps between preschooler and non-preschooler by the total sample of EGRA 2010 and EGRA 2016 (2.8-(1) and 2.8-(2)); urban sample (2.8-(3) and 2.8-(4)); and rural sample (2.8-(5) and 2.8-(6)). Across all three groups, students who attended preschool outperformed those who didn't attend by a larger margin in 2016 (post-reform) than in 2010 (pre-reform). All data points for the 2016 cohort lie farther to the left of the diagonal line with a larger bulge, which denotes the more significant gap between preschooler and non-preschooler. Notably, these gaps in test score distribution are the largest among children living in urban areas, as opposed to those among children living in rural areas. When we look closely at the area under the curve, we see that it increased from 0.52 (SE=0.01) to 0.57 (SE=0.01) between 2010 and 2016 for the entire sample. Stratified by urban and rural residence, the area under the curve 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.[68]

Through the application of the PP plots, which ensure independence from the transformation of test score scales or any threshold for specific proficiency, the gap estimates presented in Figure 2.8 offer a more accurate and visual snapshot of the gap trends between 2010 and 2016. In line with the difference in *average* test scores, the difference in the *full-distribution* of test scores depending on preschool attendance increased from 2010 to 2016. Meanwhile, the achievement gaps between preschooler and non-preschooler were more pronounced among students living in urban areas than those living in rural areas. This result is contradictory to the findings from previous studies that highlight the 'equaliser' role of preschool for children from disadvantaged backgrounds (Engle et al., 2011; Magnuson & Duncan, 2017). The gap trends shown in the current analysis signal that the benefits of preschool may be larger for advantaged children than disadvantaged children, and it could have been widened further during the reform period. These findings are further explored in Chapter 3 with an assessment of the differential role preschool has played in child outcomes by urban-rural residence and by sub-groups defined by gender, parental literacy, and home learning environment.

[68] A gap trend can be expressed as a 'change in gap' or 'difference in changes'. These are equivalent in an average-based framework but not in an ordinal framework (Ho, 2009). The Figure 2.8-PP plots capture a 'change in gap' where gaps are estimated within each year and then subtracted from each other. A 'difference in changes' indicates gap trends relying on the year-to-year linking of score scales, which has been not yet been addressed by the ordinal framework.

Figure 2.8. Probability-Probability Plot



Note: The receiver operating characteristic (ROC) curve is the plot that displays the full picture of trade-off between the sensitivity (true positive rate) and (1-specificity) (false positive rate) across a series of cut-off points. *Source:* EGRA Dataset 2010, 2016, USAID

2.8.3 Research Question 2(1): Preschool Attendance and EGRA Test Scores

To explore a change in the relations between preschool attendance and students' academic outcomes, I estimated the models described in equations (2) and (5) for early grade reading test scores. The following research question was addressed, with particular interest in the trend between 2010 and 2016: *Does the early learning reform (or large-scale expansion of preschool) strengthen or weaken the role of preschool attendance in predicting second- and third-grade students' reading achievement, as measured by EGRA test scores?* In the current analysis, the regression models were run separately by two cohorts from EGRA 2010 and EGRA 2016, and the survey weights provided with each round have been applied to ensure regional representativeness.

Table 2.11 shows the estimates of the association between preschool attendance and oral reading fluency, one key measure for reading skill acquisition among the EGRA sub-tasks. Models 1 and 2 display estimates of a parsimonious specification which only controlled for five regional and grade dummies (column 1), and age, gender, and urban-rural location (column 2).[69] In Model 1, students who attended preschool read 2.5 and 5.4 more correct words per minute, respectively, than their peers who didn't attend preschool in the 2010 and 2016 cohorts. In Model 2, once age, gender, and urban-rural locations were introduced, the difference between preschoolers and non-preschoolers significantly declined in the 2010 cohort but remained similar for Model 1 in the 2016 cohort. Model 3 added controls for household characteristics to the regressions, including the same language spoken at home and school, having reading materials at home, and father's and mother's literacy. Consistent with the idea of positive selection into pre-primary education, controlling for household characteristics reduced the coefficients between Model 2 and Model 3. The difference between preschoolers and non-preschoolers remained not significant in the 2010 cohort, whereas students in the 2016 cohort who attended preschool read approximately 4.2 more words per minute than their peers, which was statistically significant at the 0.01 level.[70]

As previously discussed, the results presented in Models 1, 2, and 3 show the associations between preschool attendance and early grade reading performance. This is unlikely to reflect

[69] Grades 2 and 3 were included as dummy variables in the model. The analysis by each grade is available upon request.
[70] In all models, coefficients were interpreted relative to a group which includes all children who experienced no preschool in the year prior to primary school. This reference group most likely includes children who were with their parents and experienced no regular out-of-home care, then the estimates would capture the substitution of preschool for home.

the causal effect of preschool attendance, due to the selection and omitted variable bias. As an intermediate step to mitigate this problem, Model 4 included school fixed effects, which accounted for unobserved heterogeneity between schools that derived from possible time-invariant factors.[71] By introducing school fixed effects, the role of preschool attendance was further reduced in the 2016 cohort, although it was still statistically significant at the 0.01 level.[72] After the massive expansion of preschool, within-school differences indicated that students who attended preschool read 2.5 more words per minute than their non-preschool peers in 2016. However, interpreting the school fixed effects estimates applied to the 2010 cohort required extra caution when there was little variation in preschool attendance and student outcomes within schools. To illustrate, in half of the primary schools in 2010, less than 10 percent of students were exposed to any form of preschool, whereas less than one-fifth of primary schools in 2016 showed this low level of preschool participation. As explained, if there is little within-subject variability, the standard errors from the school fixed effects models may be too large to tolerate, which results in some estimates that are far from the true effect (Allison, 2009; Clarke et al., 2015). Hence, in the subsequent analysis by regions/languages and by EGRA sub-tasks, the school fixed effects estimates are presented only for the 2016 cohort.

In terms of the magnitude of changes in oral reading fluency scores, as measured by correct words per minute (cwpm), the reading programme increased globally by about 6.1 cwpm on average, which was regarded as the equivalent of nearly half a year of additional schooling (Piper, Sitabkhan, Mejia, & Betts, 2018). Given that reading tests should consider the regional contexts and linguistic characteristics of the country, further exploration was pursued to see whether this global norm corresponded to the benchmark of reading proficiency in Ethiopia by regions and languages.

[71] The school fixed effects model still cannot rule out, for example, the difference between and within households. Thus, the estimates from school fixed effects remain as the 'association' between preschool attendance and student outcomes.
[72] A Hausman test (Hausman, 1978) was conducted to guide the choice between random versus fixed effects model. The results from EGRA 2010 (chi2(9) =53.4, Prob>chi2 = 0.000) and EGRA 2016 (chi2(9) =101.0, Prob>chi2 = 0.000) indicate that the null hypothesis can be rejected and suggest using the fixed effect model to calculate the regression coefficients.

Table 2.11. Preschool Attendance and Oral Reading Fluency

Oral Reading Fluency (correct words per minute, cwpm)	(1)	(2)	(3)	(4)
	EGRA 2010 Cohort			
Attended preschool	2.46**	1.24	0.46	-0.39
(SE)	(1.22)	(1.44)	(1.55)	(1.94)
R-Squared	0.15	0.18	0.20	0.10
Observations (schools)	9,121	9,121	9,121	9,121 (237)
	EGRA 2016 Cohort			
Attended preschool	5.36***	5.45***	4.15***	2.48***
(SE)	(0.86)	(0.87)	(0.78)	(0.61)
R-Squared	0.22	0.24	0.26	0.11
Observations (schools)	8,332	8,332	8,332	8,332 (225)
	Controls included in EGRA 2010 and 2016 Cohort			
Region and grade dummies	Yes	Yes	Yes	Yes (Grade only)
Age, gender, location	No	Yes	Yes	Yes
Household characteristics	No	No	Yes	Yes
School fixed effects	No	No	No	Yes

Note: (1) Models 1, 2, and 3 account for controls as indicated and include sampling weight; (2) Model 4 uses school fixed effects and includes sampling weight; Number of schools in parentheses; (3) EGRA 2010: linearised standard errors (from svy command) in parentheses; (4) EGRA 2016: robust standard errors, clustered at school level, in parentheses; (5) Appendix Tables D.1. (EGRA 2010) and D.2. (EGRA 2016) present the full regression results.
*** p<0.01, ** p<0.05, *p<0.1
Source: EGRA Dataset 2010, 2016, USAID

Table 2.12 presents the regression results of preschool attendance and oral reading fluency by regions (language for SNNP), with reference to the benchmark for Ethiopia's reading proficiency levels. The first column shows the difference in the highest reading proficiency benchmark (Level 4: reading with full comprehension) between Grade 2 and Grade 3. Specifically, this denotes the expected level of improvement in oral reading fluency when students progressed from Grade 2 to Grade 3.[73] As previously described, the benchmark for reading proficiency level in Ethiopia was developed distinctively, based on each language-specific metric; thus, the grade differences in the Level 4 benchmark varied by language from 5 cwpm in Somali to 10 cwpm in Amhara and Oromia. For example, in Amhara and Oromia, where the benchmark increased 10 cwpm from Grade 2 to Grade 3, one can regard 'reading 10 more correct words per minute' as the equivalent of one year of additional schooling.

Although the 2010 cohort didn't show any significant difference in average score, a positive association was observed in Amhara in the OLS model with full control variables (Model 1). For the 2016 cohort, the results from both the OLS with full control variables and school fixed

[73] Globally, the benchmark ranges between 45 and 60 words per minute (Kelly & Graham, 2017), but Ethiopia's case shows more various ranges by mother tongue languages.

effects were presented; all associations across regions were statistically significant, which varied from 3 to 6 cwpm, except Tigray (Model 2). When school fixed effects were introduced to the 2016 cohort (Model 3), the results of three regions (Amhara, Oromia, and Somali) remained significant, whereas the results of the SNNP region became insignificant. In this preferable specification, students in Amhara and Oromia read 2.6 and 3.6 more words per minute (significant at 0.05 levels), respectively, which was nearly the same as one-third of a single year of schooling. Most interestingly, in Somali, where 5 cwpm was regarded as the equivalent of one year of additional schooling, the difference between preschoolers and non-preschoolers was 5.0 cwpm in the school fixed effects specification. In other words, the preschool advantage in the Somali region was relatively comparable to the learning that occurred in one academic year.

To check the robustness of the regression results, the results from *kernel*-based propensity score matching (PSM) are presented in Models 4 and 5 (more details about PSM in Appendix A).[74] By controlling the same observed variables to construct a matching group, the results of PSM were similar or slightly larger than estimates of the school fixed effects model. Given that Somali is one of the historically disadvantaged regions in Ethiopia, these relatively encouraging results indicate that the positive contribution of preschool attendance was more significant in the marginalised areas.

Table 2.12. Preschool Attendance and Oral Reading Fluency by Regions

Region	Grade difference in benchmark 'Level 4' (Grade 3 benchmark–Grade 2 benchmark)		(1) 2010 OLS	(2) 2016 OLS	(3) 2016 School Fixed-Effects	(4) 2010 Propensity Score Matching	(5) 2016 Propensity Score Matching
Average			0.46 (1.55)	4.15 (0.78) ***	2.48 (0.61) ***	1.98**	3.09***
Tigray	7 cwpm	(62-55)	1.99 (1.15)	1.41 (1.07)	0.28 (0.88)	4.94***	0.35
Amhara	10 cwpm	(60-50)	4.81 (2.04) **	3.40 (1.64) **	2.59 (1.31) **	4.82**	4.57***
Oromia	10 cwpm	(58-48)	-1.00 (4.43)	5.13 (1.80) ***	3.58 (1.57) **	2.11	4.09***
Somali	5 cwpm	(55-50)	2.60 (2.09)	5.86 (1.99) ***	5.08 (1.65) ***	-1.77	4.80***
SNNP	8 cwpm	(53-45)	-2.07 (1.00) *	3.05 (1.49) **	0.91(1.15)	0.02	1.56*

Note: (1) Grade difference in Level 4 benchmark is calculated as follows: (G3 Level 4 threshold - G2 Level 4 threshold); (2) The same set of covariates used across all models; (3) Average estimates in Models 1 to 2 correspond to Table 2.11-Model 3 and include sampling weight; (4) Model 3 uses school fixed effects and includes sampling weight; (5) EGRA 2010: linearised standard errors (from svy command) in parentheses; (6) EGRA 2016: robust standard errors, clustered at school level, in parentheses.
*** p<0.01, ** p<0.05, *p<0.1 *Source:* EGRA Dataset 2010, 2016, USAID

[74] PSM assumes conditional independence, which requires rich data with child development history and institutional information. Considering the lack of richness in school-based EGRA information compared to another household survey, I decided to use PSM as the robustness check instead of the primary approach for the present study.

In addition to the results by region and language, I analysed the association between preschool attendance and test scores from other EGRA sub-tasks: letter sounds, familiar words, and invented words recognition, reading and listening comprehension. Table 2.13 summarises the results from these sub-tasks in relation to preschool attendance (oral reading fluency presented again for comparison). It reports that, in 2010, only two sub-tasks—letter sounds and invented words recognition—showed small but positive associations with preschool attendance in Model 1 (OLS). In 2016, the associations were larger than in 2010 and became positive across all sub-tasks of EGRA, which were statistically significant at the 0.01 levels in Model 2 (OLS). In Model 3, with the school fixed effects specification for the 2016 cohort, the gains from preschool attendance were reduced yet remained statistically significant ($p < 0.01$), except listening comprehension. In terms of the magnitude of association, the gap between preschoolers and non-preschoolers was the largest in letter sounds recognition and the smallest in listening comprehension for both cohorts.

Table 2.13. Association between Preschool Attendance and Early Grade Reading Outcomes, 2010 and 2016

EGRA Task	(1) 2010 OLS		(2) 2016 OLS		(3) 2016 School-Fixed	
	Coef. (SE)	Effect Size (SE)	Coef. (SE)	Effect Size (SE)	Coef. (SE)	Effect Size (SE)
Oral reading fluency (cwpm)	0.46 (1.55)	0.02 (0.07)	4.15*** (0.78)	0.20*** (0.03)	2.48*** (0.61)	0.12*** (0.03)
Letter sounds (clpm)	3.33** (1.53)	0.11** (0.05)	6.48*** (0.91)	0.20*** (0.03)	3.46*** (0.87)	0.11*** (0.03)
Familiar words (cwpm)	1.46 (1.38)	0.08 (0.07)	4.25*** (0.60)	0.21*** (0.03)	2.28*** (0.59)	0.11*** (0.03)
Invented words (cwpm)	1.61** (0.80)	0.11** (0.06)	2.84*** (0.46)	0.19*** (0.03)	1.60*** (0.43)	0.11*** (0.03)
Reading comprehension (% of correct answer)	1.34 (1.92)	0.05 (0.07)	4.67*** (0.85)	0.17*** (0.03)	2.78*** (0.82)	0.10*** (0.03)
Listening comprehension (% of correct answer)	0.21 (1.64)	0.01 (0.06)	1.85** (0.76)	0.07** (0.03)	1.20 (0.77)	0.04 (0.03)
Observation	9,121	9,121	8,332	8,332	8,332	8,332

Note: (1) The same set of covariates was used across all models; (2) Models 1 and 2 include sampling weight; (3) Model 3 uses school fixed effects and includes sampling weight; (4) EGRA 2010: linearised standard errors (from svy command) in parentheses; (5) EGRA 2016: robust standard errors in parentheses.
*** p<0.01, ** p<0.05, *p<0.1
Source: EGRA Dataset 2010, 2016, USAID

To simplify the interpretation of the results, Table 2.13 also presents the effect size of each model. The effect size is the difference in standard deviation (SD) using the standardised scores (z-score) of the EGRA sub-tasks within the sample. In 2010, recognition of letter sounds and invented words presented a small effects size at 0.11 SD ($p < 0.05$) in Model 1. In 2016, five sub-tasks (except listening comprehension) showed effect sizes close to 0.20 SD in Model 2 and 0.11 SD in Model 3 (all estimates are $p < 0.01$). These effect sizes are considered small (Cohen, 1992) but consistent with the small-to-moderate effect size of educational interventions in other LMICs (McEwan, 2015). In a large-scale preschool expansion in Argentina that created exogenous variations between cohorts and provinces, Berlinski et al. (2008) found that the impact of an additional preschool place per child was 0.23 SD ($p < 0.01$) on average for students' learning outcomes in Grade 3. In the Sub-Saharan Africa context, among students aged 6-16 in Kenya and Tanzania, Bietenbeck et al. (2017) revealed that students who participated in preschool scored about 0.10 SD ($p < 0.01$) higher on standardised cognitive tests than peers who didn't participate, when applying the household (sibling) fixed effects to control for the unobserved, time-invariant variables between households.

In the analysis above, I included grade as the dummy variable (i.e., grade fixed effects). When looking at how the influence of preschools differed by grade, the benefits of preschool were more pronounced in Grade 3 than in Grade 2 in 2010.[75] This may be counter-intuitive, given that, in many studies using longitudinal data (e.g., Bassok et al., 2018; Magnuson et al., 2007), the initial link of preschool with cognitive outcomes tended to be diminished along with the grade progression of a child, so-called 'fade-out' effects. However, in Ethiopia, Woldehanna and Gebremedhin (2012) found that the positive influence of preschool on children's receptive vocabulary and early math skills were larger at age 8 than at age 5. Similarly, Biethentek et al. (2017) found a more significant contribution of preschool for an older cohort (10-12, 13-16 years old) than a younger cohort (7-9 years old) in Kenya. While the achievement gaps depending on preschool attendance were relatively stable and persistent in the 2016 cohort, the widening gaps by grades in the 2010 cohort need further investigation.

2.8.4 Research Question 2(2): Preschool Attendance and the Probability of Non-Reader and Proficient Reader

I continued to explore the relationship between preschool attendance and early grade reading performance, as measured by the probability of being a non-reader and a proficient reader, two frequently used indicators in the global EGRA administration. The following research question was addressed: *Does the early learning reform (or large-scale expansion of preschool) strengthen or weaken the role of preschool attendance in predicting second- and third-grade*

[75] The results from analysis by Grade 2 and Grade 3 are available upon request.

students' reading achievement, as measured by the probability of being a non-reader and a proficient reader? I illustrate the results of a multivariate logistic regression model as expressed by odds ratio, marginal effects, and predicted probabilities.

Odds ratio. Table 2.14 presents odds ratios regarding the relationship between preschool attendance and students' early grade reading performance. The first three columns of each outcome variable display the results from logistic regression, after controlling for region and grade dummies (Models 1 and 5), age, gender, and urban-rural location (Models 2 and 6), and household characteristics (Models 3 and 7). The last column introduces the school fixed effects (Models 4 and 8), which captured the achievement gaps within schools as a function of preschool attendance. In the 2010 cohort, the odds ratio for non-reader between preschoolers and non-preschoolers was statistically not significant across the models. In contrast, the results of the 2016 cohort indicated that, when children attended preschool, the odds of being a non-reader decreased by 51 percent, holding all other variables constant (Model 3). When looking at the within-school difference, preschoolers were 23 percent less likely to be non-readers in 2016 (Model 4).

With regard to the likelihood of being a proficient reader, it was continuously not significant in the 2010 cohort. In contrast, preschoolers in 2016 were 38 percent more likely to be a proficient reader than non-preschoolers (Model 7), although this difference declined significantly in the school fixed effects specification (Model 8). A significant drop in the school fixed effects model implied that, without controlling for unobserved, time-invariant differences between schools, the estimates from the OLS or random-effect model were likely to be biased (Clarke et al., 2015). This perhaps justifies the application of the school fixed effects approach used in the current analysis. Regarding the observed characteristics of primary school (post-preschool inputs in the present study), Chapter 3 further investigates the role of subsequent schooling experiences that potentially mediate the link of preschool attendance with early grade reading performance.

Lastly, recall that the reference categories of non-reader and proficient reader are not mutually exclusive. The non-reader group was compared against all three other categories in reading proficiency (students with limited, increasing, or full reading fluency), which included proficient readers from two upper categories. Similarly, the proficient reader group was compared against two lower categories in reading proficiency, which included the non-reader group.

Table 2.14. EGRA 2010 and EGRA 2016: Non-Reader and Proficient Reader

	Non-Reader (*Logit*)				Proficient Reader (*Logit*)			
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
	EGRA 2010 Cohort							
Preschool	0.94	0.99	1.06	1.28***	1.23*	1.11	1.05	0.91
(SE)	(0.13)	(0.16)	(0.17)	(0.04)	(0.13)	(0.11)	(0.10)	(0.21)
Pseudo R2	0.10	0.11	0.13	-	0.07	0.08	0.09	-
Observations	9,121	9,121	9,121	8,812 (229)	9,121	9,121	9,121	9,121
	EGRA 2016 Cohort							
Preschool	0.46***	0.42***	0.49***	0.77***	1.57***	1.58***	1.38***	1.07***
(SE)	(0.05)	(0.04)	(0.05)	(0.01)	(0.15)	(0.15)	(0.12)	(0.01)
Pseudo R2	0.16	0.18	0.20	-	0.08	0.10	0.11	-
Observations	8,332	8,332	8,332	7,461 (202)	8,332	8,332	8,332	8,020 (214)
	Controls included in EGRA 2010 and EGRA 2016 Cohort							
Region/Grade	Yes	Yes	Yes	Grade only	Yes	Yes	Yes	Grade only
Age, gender, location	No	Yes	Yes	Yes	No	Yes	Yes	Yes
Household Character	No	No	Yes	Yes	No	No	Yes	Yes
School fixed effects	No	No	No	Yes	No	No	No	Yes

Note: (1) All Models 1 to 8 account for controls as indicated and include sampling weight; (2) Model 4 and Model 8 use school fixed effects; Number of schools in parentheses; (3) EGRA 2010: linearised standard errors (from *svy* command) in parentheses; (4) EGRA 2016: robust standard errors, clustered at school level, in parentheses; (5) In Model 4 from EGRA 2010, the sample of 8 groups (309 obs.) in non-reader and 10 groups (390 obs.) in proficient reader dropped as multiple positive or negative outcomes within groups encountered; (6) In Model 8 from EGRA 2016, the sample of 23 groups (871 obs.) in non-reader and 11 groups (312 obs.) dropped for the same reason. (7) Appendix Tables D.3. to D.6 present the full regression results.
*** $p<0.01$, ** $p<0.05$, * $p<0.1$
Source: EGRA Dataset 2010, 2016, USAID

***Marginal effects and predicted probabilities*.** Table 2.15 presents the AMEs of preschool attendance on non-reader and proficient reader and the estimates stratified by regions.[76] Recall that the *average* marginal effects were computed using some fixed values (i.e., preschool and regions) and observed values for all other variables, which presented the averaged marginal effects of each predictive value. Regarding the probability of being a non-reader, having attended preschool decreased a child's probability of being a non-reader by 11 percentage points on average ($p < 0.01$) in the 2016 cohort; there was no significant discrete change in the 2010 cohort. In four out of five regions (Tigray, Oromia, Somali, and SNNP), preschool attendance reduced the probability of being a non-reader significantly, from 10 percentage points ($p < 0.01$) in Tigray to 15 percentage points ($p < 0.01$) in Somali. In Amhara, the highest performing region in early grade reading, the marginal effect of the association between preschool and students' outcomes was much smaller but statistically significant, at 3 percentage points ($p < 0.01$).

[76] For factor variables used in the current paper, that is, preschool attendance, marginal effects can be interpreted as *discrete* change when a regressor changes by a fixed amount from 0 (non-preschool) to 1 (preschool).

When it comes to the probability of being a proficient reader, similar patterns were observed in 2010 and 2016. On average, the 2010 cohort didn't show any significant gains from preschool attendance, whereas in the 2016 cohort, having attended preschool increased a child's probability of being a proficient reader by 7 percentage points ($p < 0.01$). The marginal effects of this association were similar across regions, from 5 percentage points in Somali to 7 percentage points in Tigray, Amhara, and SNNP.

In addition, I found no significant correlation between the marginal effects of preschool attendance and the overall preschool participation rate by region. For example, Somali (highest gains in reducing non-readers) and Amhara (least gains in reducing non-readers) showed similar trends in preschool participation; in 2010, less than 10 percent of students attended preschool in two regions, which increased to 27.2 percent in Somali and 22 percent in Amhara in 2016, after the rapid expansion of public preschool. In SNNP, the marginal effect of the association between preschool and student outcomes was about average or slightly above other regions, while the preschool participation rate there increased most rapidly, from 18.2 percent to 64.4 percent between 2010 and 2016.

Table 2.15. Average Marginal Effects of Preschool Attendance, by Region

EGRA 2010 Cohort						
	(1)	(2)	(3)	(4)	(5)	(6)
Non-Reader	**Average**	**Tigray**	**Amhara**	**Oromia**	**Somali**	**SNNP**
All grades	0.01	0.01	0.01	0.01	0.01	0.01
(SE)	(0.03)	(0.02)	(0.03)	(0.03)	(0.03)	(0.04)
EGRA 2016 Cohort						
Non-Reader	**Average**	**Tigray**	**Amhara**	**Oromia**	**Somali**	**SNNP**
All grades	-0.11***	-0.10***	-0.03***	-0.13***	-0.15***	-0.14***
(SE)	(0.01)	(0.01)	(0.00)	(0.01)	(0.02)	(0.02)
EGRA 2010 Cohort						
Proficient Reader	**Average**	**Tigray**	**Amhara**	**Oromia**	**Somali**	**SNNP**
All grades	0.01	0.01	0.01	0.01	0.01	0.01
(SE)	(0.02)	(0.02)	(0.02)	(0.02)	(0.02)	(0.01)
EGRA 2016 Cohort						
Proficient Reader	**Average**	**Tigray**	**Amhara**	**Oromia**	**Somali**	**SNNP**
All grades	0.07***	0.07***	0.07***	0.06***	0.05***	0.07***
(SE)	(0.02)	(0.02)	(0.02)	(0.01)	(0.01)	(0.02)

Note: (1) The table shows marginal effects of the association between preschool attendance and the probability of being a non-reader/proficient reader by average (column 1) and stratified by 5 regions (columns 2 to 6). (2) All estimates are based on a logit regression model including sampling weight; (3) Robust standard errors in parentheses. *** p<0.01, ** p<0.05, *p<0.1
Source: EGRA Dataset 2010, 2016, USAID

Figure 2.9 offers a visual representation of the predicted probabilities of being a non-reader as a function of preschool attendance, stratified by region. While marginal effects captured the magnitude of the achievement gaps between preschoolers and non-preschoolers, the visualised probabilities depicted how these gaps changed dynamically between 2010 and 2016 across the five regions. As presented in Figure 2.9, each region showed a different pattern of improving or deteriorating early grade reading performance as a function of preschool attendance. Looking into regional variations and patterns was particularly pertinent to the Ethiopian context, where the large regional imbalance in students' academic achievement prevailed and persisted across different levels of education. As it started from early grade reading in Grades 2 and 3 in the present study, the regional disparities in reading proficiency level have been documented at Grade 4, measured by National Learning Assessment (World Bank, 2016), and language proficiency level at Grade 12, measured by National Standardised Examinations (Tesema & Braeken, 2018).

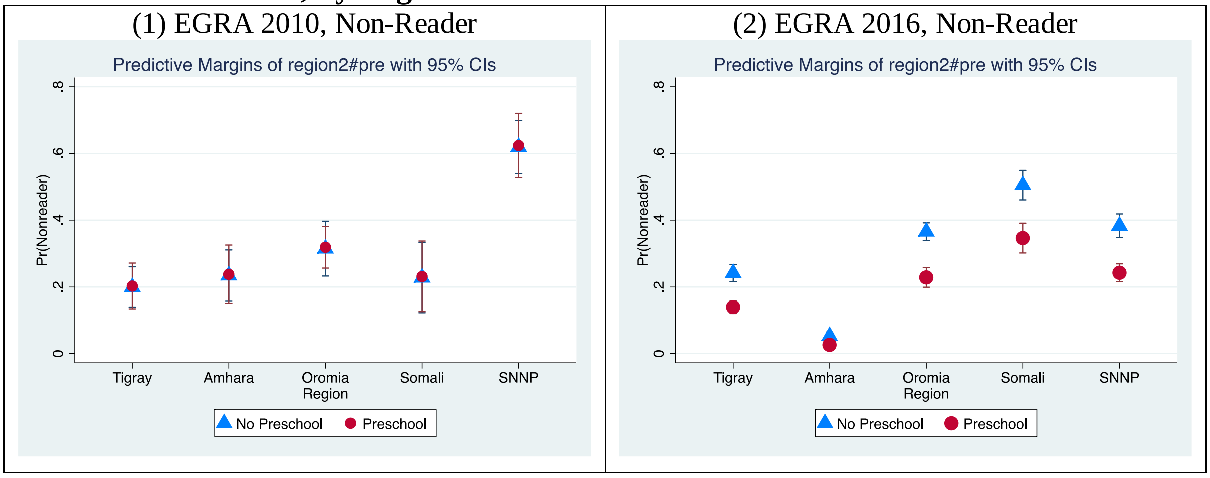
Four out of the five regions, Tigray, Amhara, Oromia, and SNNP, showed an overall decline in the proportion of non-readers between 2010 and 2016, while the magnitude of this decline (positive gains) was much higher among preschoolers than non-preschoolers. For example, in SNNP, the most significant decline of the non-reader share was observed between 2010 and 2016, which favoured preschoolers. Among preschoolers in SNNP, the non-reader share decreased from 62 percent to 24 percent between 2010 and 2016 (38 percentage points), while this share among non-preschoolers decreased from 62 percent to 38 percent (24 percentage points) during the same period. Amhara, which had the smallest share of non-readers, also reported a significant decline in the non-reader share among both preschoolers and non-preschoolers, at 21 and 18 percentage points difference, respectively.

The achievement gaps associated with preschool attendance widened between 2010 and 2016 in two regions: in Tigray, the non-reader share among preschoolers decreased 6 percentage points (from 20% to 14%), while this share among non-preschoolers increased 4 percentage points (from 20% to 24%). A similar pattern was observed in Oromia, where preschoolers were less likely to be a non-reader (9 percentage points) and non-preschoolers, with an absence of early learning experience before entering primary school, were more likely to be a non-reader (6 percentage points). As an exception, in Somali there was a sharp increase in the overall non-reader share between 2010 and 2016.[77] During this period, the non-reader share among

[77] This is partly explained by the convenience sampling applied to Somali specifically in 2010, due to the political instability during the election period. The convenience sampling was not the case in 2016.

preschoolers increased 12 percentage points (from 23% to 35%); it showed a steeper increase among non-preschoolers, at 28 percentage points (from 22% to 50%).

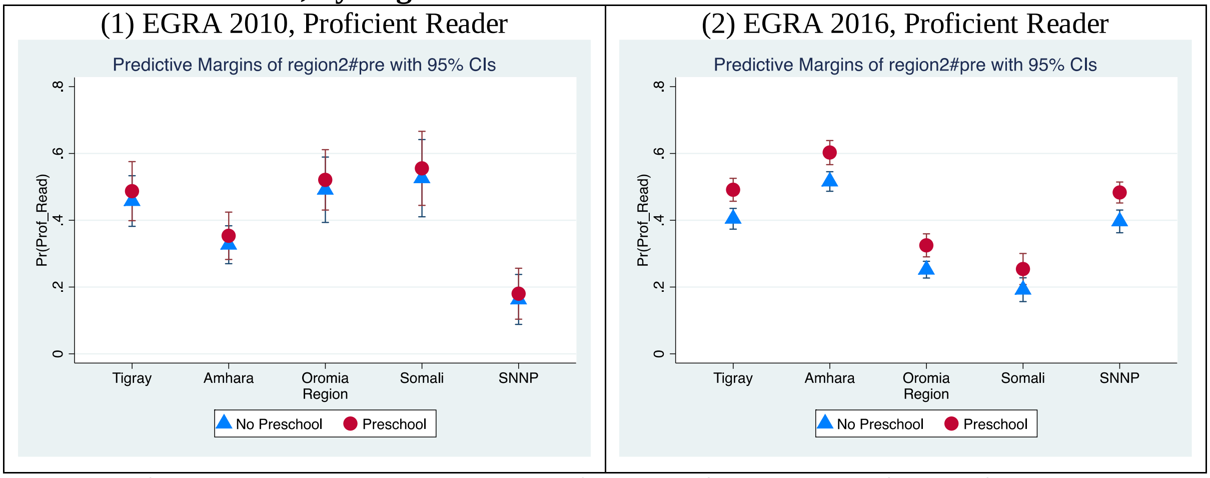
Figure 2.9. Average Predicted Probability of Being a Non-Reader as a Function of Preschool Attendance, by Region



Note: (1) The figure shows average predicted probability of non-reader as a function of preschool attendance, conditioned on the location (five regions in Ethiopia); (2) All estimates are based on a logit regression model including sampling weight; (3) The lines represent the upper and lower bounds of the confidence interval. Confidence level = 0.95.
Source: EGRA Dataset 2010, 2016, USAID

When it comes to the probability of being a proficient reader, as presented in Figure 2.10, more mixed patterns were observed across the five regions, yet the magnitude of positive gains was continually larger for preschoolers than non-preschoolers. In Amhara and SNNP, there was a significant increase in the probability of being a proficient reader, which favoured preschoolers. In Amhara, the share of proficient readers increased from 35 percent to 60 percent among preschoolers (25 percentage points) and from 33 percent to 52 percent among non-preschoolers (19 percentage points). Similarly, in SNNP there was a sharp increase in the share of proficient readers for pre-schoolers, from 18 percent to 48 percent (30 percentage points), as well as for non-preschoolers, from 16 percent to 40 percent (24 percentage points). In Tigray, about half of preschoolers were proficient readers in both 2010 and 2016, but this share decreased only among non-preschoolers, from 46 percent to 40 percent. Alarmingly, Oromia and Somali showed deteriorating trends between 2010 and 2016. In Oromia, the probability of being a proficient reader dropped from 52 percent to 32 percent among preschoolers (20 percentage points), and from 49 percent to 25 percent among non-preschoolers (24 percentage points). The largest decline was observed in Somali, where the proficient reader share dropped to 31 percentage points among preschoolers (from 56% to 25%) and 34 percentage points among non-preschoolers (from 53% to 19%) between 2010 and 2016.

Figure 2.10. Average Predicted Probability of Being a Proficient Reader as a Function of Preschool Attendance, by Region



Note: (1) The figure shows average predicted probability of being a proficient reader as a function of preschool attendance, conditioned on the location (five regions in Ethiopia); (2) All estimates are based on a logit regression model including sampling weight; (3) The lines represent the upper and lower bounds of the confidence interval. Confidence level = 0.95.
Source: EGRA Dataset 2010, 2016, USAID

Taken together, the results in Table 2.15 and Figures 2.9 and 2.10 show an interesting pattern of how preschool attendance was associated with students' learning outcomes during a massive expansion of public preschool in Ethiopia. The wider access to preschool was accompanied by meaningful gains in early grade reading performance, which were substantiated by the lower chance of becoming non-readers and higher chance of becoming proficient readers. However, further exploration of the dynamic patterns across regions revealed that the achievement gap between preschoolers and non-preschoolers was widening, with a huge regional imbalance in student achievement.

2.9 Discussion

By exploiting the massive expansion of public preschool (O-Class) in Ethiopia, the present study addresses gaps in the previous literature related to the influence of large-scale ECE programmes in the LMIC context. Specifically, I examine the *trends* of the relationships between preschool attendance and students' early grade reading performance between 2010 and 2016, when enrolment rates in pre-primary soared by nearly ten times. I found that positive and statistically significant relationships were observed only after the expansion of public preschool (2016) when comparing the period before the expansion (2010).

This is consistent with previous studies reporting the small to moderate effects of a large-scale expansion of pre-primary education on students' performance in reading and mathematics in Grade 3 in Argentina (Berlinski et al., 2009), and on children's school readiness in language, cognitive, and socio-emotional development after long exposure to preschool (three years) in rural Indonesia (Brinkman et al., 2017). My findings are also consistent with prior evidence on ECE participation in Sub-Saharan Africa using a nationally representative sample: short-term benefits of preschool on 6-year-old children's school readiness in multiple cognitive and non-cognitive domains in Zambia (McCoy, Zuilkowski, et al., 2017); small gains in language test scores from Grade 1 to Grade 6 for students in South Africa (Berg et al., 2013); and better performance on the literacy and numeracy tests sustained across age groups from 7- to 16-year-olds in Kenya and Tanzania (Bietenbeck et al., 2017). I acknowledge that the influences (effect sizes) of preschool attendance in the present study, as well as the previous studies, are considered small (Cohen, 1992) but statistically significant across highly diverse settings. This is consistent with the claim that the estimated effects of large-scale ECE programmes are generally smaller than those found for small-scale programmes (Duncan & Magnuson, 2013).

Previous studies in Ethiopia have shown the benefits of preschool attendance on cognitive development at age 8 (Woldehanna & Gebremedhin, 2012) and secondary school completion (Woldehanna & Araya, 2017) among urban children; however, there is no such evidence after the massive expansion of ECE following the enactment of the early learning reform in 2010. Rather, my findings among the EGRA 2010 cohort (before the expansion) differ from those reported by Woldehanna and Gebremadin (2012), who studied seemingly compatible preschool attendees in the pre-reform context in Ethiopia.[78] Their findings point to an effect size of 0.36 on a receptive vocabulary test of preschool attendees versus non-attendees (Grade 2 equivalent), whereas the present study found no significant association between attending preschool and reading achievement at Grade 2 for the EGRA 2010 cohort. This may be attributed in part to the different sampling frame: sample students from the previous study were restricted to those living in urban areas, including Addis Ababa (the area excluded in the current study), and they were not a regionally representative sample (Outes-Leon & Sanchez, 2008). Another reason for the inconsistency may stem from the quality of preschool services provided (e.g., kindergarten vs. religion-based preschool); however, there is not sufficient information to examine the quality aspects of the preschools where children attended.

[78] The EGRA 2010 and the Young Lives samples attended preschool between 2005 and 2008. Between the two samples, four regions are overlapped from Amhara, Oromia, SNNP, and Tigray, then the only difference is Addis Ababa in Young Lives and Somali in EGRA 2010.

As emerging evidence, my findings from the early learning reform in Ethiopia contribute to our understanding of whether the benefits of ECE from small-scale trials can be replicated when ECE is implemented on a larger scale, especially in low-resource settings. The previous studies found a positive link between access to preschool and students' learning outcomes in Sub-Saharan Africa (e.g., Biethentek et al., 2017; Berg et al., 2013), but the maturity of the ECE system and its coverage is not commensurate with that in Ethiopia. For example, among the nationally representative study sample, enrolment in preschool increased from 79 percent to 83 percent between 1997 and 2004 in Kenya, and from 64 percent to 72 percent between 2007 and 2008 in South Africa.[79] Caution is required in extrapolating previous findings to poorly resourced settings or the early period of the system expansion. Average gross enrolment ratios in pre-primary remain at 20 percent and 36 percent in low- and lower-middle income countries, respectively. Reflecting on the scale-up pathways suggested by Yoshikawa et al., (2018) and Marmot (2010), existing evidence has focused on the 'big to better' process, whereas the present study was the first to add evidence from a 'small (nothing) to bigger' pathway operating at a national scale. This may be more relevant evidence for the policymakers, practitioners, and researchers working in LMICs who envisage universal pre-primary education, which was recently elevated to the top priority on the national and international policy agenda.

Through the lens of bioecological theory (Bronfenbrenner, 1979; Bronfenbrenner & Morris, 2006), the present study attempts to fill the gap in the literature of processes linking policy variation to child developmental variation (Yoshikawa & Hsueh, 2001). I looked not only into a shift in the ECE landscape as a change of the macrosystem children and families inhabit, but also into a shift in the test score distribution of *all* children in the sample. This reaffirmed that the early learning reform in Ethiopia induced a big stride forward in the education system, from being elite-oriented to including a wider representation of society. Previous studies were limited to examining *mean* differences in child outcomes across policy environments, whereas I employed the novel approach of using PP plots with students' test score *distribution* (Ho, 2009; Ho & Reardon, 2012), which depicted a change in the achievement gap based on preschool attendance on a robust scale. The gap trends also suggested that the benefits of preschool can be more pronounced in students living in urban areas than those in rural areas,

[79] This figure is retrieved from UNESCO's UIS database (January 2019). This can be different from the national education statistics.

which was relevant to the equity concerns raised by the previous study. In South Africa, Berg et al. (2013) found that preschool attendance had virtually no measurable influence on students' academic achievement in the schools from the lower wealth quintile, while there was a more discernible influence for the highest quintile schools (further investigation presented in Chapter 3). Although high value was placed on monitoring trends in preschool participation related to learning outcomes in order to better inform ECE policy and service provision (Bassok et al., 2018; Connor et al., 2016), this area has been overlooked in many studies conducted in LMICs.

This paper also adds to the evidence on the wide disparities in academic achievement within Ethiopia. As each region forged ahead with its own plans for expansion, the magnitude of preschool benefits varied considerably by region.[80] This point is relevant to the previous studies in Ethiopia that highlighted substantial regional differences in educational opportunities and academic achievement at Grade 4 (NEAEA, 2016) and Grade 12 (Tesema & Braeken, 2018). Importantly, in the present study, while regions experienced an overall decline in the share of non-readers, these gains always favoured preschool attendees to a greater extent than non-attendees. For example, in SNNP (Sidamu)—the region that experienced the most dramatic increase in preschool access—the share of non-readers among preschoolers decreased by 41 percentage points (from 62% to 21%) while it decreased among non-preschoolers by 27 percentage points (from 62% to 35%). A huge regional imbalance and growing inequalities imply that government efforts toward universal pre-primary need to focus on promoting equitable access to high-quality ECE opportunities between and within regions.

Another interesting result can be found in the largest gains from preschool in Somali, one of the historically disadvantaged regions (so-called emerging regions) in Ethiopia. In this region, after the large-scale expansion of O-Class in 2016, the benefits of preschool on early grade reading performance was almost equivalent to a single year of additional schooling, even after accounting for the between-school difference. Preschool attendance was also associated with a 15 percent lower chance of becoming a non-reader, which is above the national average (10%). This finding aligns with previous studies that emphasised the compensatory role of preschool for children from more disadvantaged background (Brinkman et al., 2017); however, further

[80] Meanwhile, there is little correlation between regions' preschool expansion rates and preschool benefits in students' academic achievement.

investigation is needed to explain this finding relative to the overall academic achievement trends in this region.[81]

While positive influences on child outcomes after the mass expansion of preschool are encouraging, why was there a 'stronger' association between preschool and academic achievement for the post-reform period but not for the pre-reform period? Initially, I hypothesised a 'small' association in the recent cohort (2016) with respect to a sudden influx into the system of many previously excluded young children. This is, in fact, what has been witnessed since the early 1990s during the universalisation of primary education. In Ethiopia, this movement shows that access-oriented policy reform could lead to the deterioration of learning outcomes, given the inclusion of many young people from deprived backgrounds (Dom, 2010).

My hypothesis also reflects quality concerns in most O-Classes nationwide, such as an unqualified workforce, a lack of an age-appropriate curriculum and learning materials, a lack of infrastructure and financial resources, and poor managerial capacity among regional and local officers, among others (Rossiter et al., 2018; Woodhead et al., 2017; Teferra & Hagos, 2016). Unlike the previous studies on the large-scale ECE expansion, which were accompanied by a strong government financial commitment (often combined with international aid) and targeted mechanisms for rural, disadvantaged populations (Bastos et al., 2017; Berlinski et al., 2009, 2008; Bouguen et al., 2014; Brinkman et al., 2017), the early learning reform in Ethiopia was rolled out without such strategies. The education sectoral budget allocated to pre-primary remained scarce, about 3 percent of the total during the initial reform period (Rossiter et al., 2018). Nevertheless, I found a reverse pattern suggesting a 'stronger' association between preschool and child outcomes after the reform, once wider access to preschool was instituted, as partly intended by the reform aiming to improve school readiness of students.

Although beyond the scope of this analysis, I extrapolated a few possible reasons for this finding. First, the current study documents gaps in academic achievement between children who attended preschool and those who did not, thus it may relate to *greater selection bias,* especially for preschool non-attendees, during the major shift in the ECE system. This transition may leave a highly disadvantaged population excluded from early learning

[81] With respect to the sampling frame, although EGRA 2010 adopted convenience sampling for Somali due to the political instability, EGRA 2016 used the same sampling procedure it used for Somali with four other regions (e.g., random sampling after excluding the UNICEF Emergency Priority Zones).

opportunities. In other words, nearly all 4- to 6-year-old children (95%) did not have an opportunity to attend preschool in 2010; however, after the massive expansion of O-Class, 50 percent of children from the marginalised communities in remote areas still do not have access to pre-primary education.

Moreover, in addition to the selection bias caused by family wealth and parental choice, preschool attendance was partially explained by supply factors—that is, better-resourced schools may have introduced O-Class earlier than schools with few resources. Provided that children who benefited most from the early learning reform were attending primary schools with relatively high educational attainment, the comparison groups that entered primary school without any preschool experience were more likely to come from the low-performing schools in underprivileged communities. Indeed, a number of recent studies have highlighted the importance of understanding the *counterfactual* (or composition effects) when investigating the impact of preschool programmes (Bouguen et al., 2014; Zhai, Brooks-Gunn, & Waldfogel, 2014). Collectively, this suggests that a larger association in the post-reform period does not necessarily imply that the rapid expansion yielded the intended benefits—an important area for future research.

Perhaps the downside of my findings from the early learning reform is that the gaps between preschool attendees and non-attendees widened, which may have exacerbated educational inequality. These findings shed light on the need for more attention to children who did not attend preschool. If we project the learning trajectories of both groups, drawing from the observed trends, the achievement gaps would be wider, due to the cumulative advantage for preschoolers and cumulative disadvantage for non-preschoolers. While improving the quality of current pre-primary service provision has been prioritised, equivalent attention should be given to targeted interventions for children who are deprived of early learning opportunities before entering primary school.

It is worth noting that there are some innovative approaches to reaching more young children in Ethiopia. The first case is community-based O-Class in SNNP (Rossiter et al., 2018). This is a local adaptation of school-based O-Class, which was directed to be established within the compounds of government primary schools. To reach the wider community, O-Class in SNNP has been expanded through another existing structure, including religious institutions, farmers' training centres, and community centres (personal interview with REB).[82] The second case is the Accelerated School Readiness programme initiated by the government with support from UNICEF. The ASR programme provides a non-formal, short-term supplementary education for 6-year-olds who did not attend preschool, as an ‘interim’ approach to bridge universal access to formal preschool provided in kindergarten or O-Class. There is a need to maintain policy attention on efforts to further reduce barriers to preschool for at-risk subpopulations, and to utilise various channels that could reach children who are still excluded from early learning opportunities.

2.10 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 were limited by the characteristics of the data I had available. 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. Also, given that several additional EGRA tests were administered in Ethiopia in 2013 and 2014, there may be some unobservable aspects, such as the improved capacity of EGRA administrators, assessors, and field workers.

Second, there is limited information on SES or the household income levels of children, especially in the 2016 EGRA data. One explanation for excluding SES measures in the survey was the validity of SES measures self-reported by 8- and 9-year-old students. Within the scope of the data available, I mitigated the absence of SES measures by using other indicators, such as urban-rural residence, father’s and mother’s literacy, and reading materials at home, yet there is a certain restriction. To overcome this, a future study could consider linking the EGRA with a household-based survey that would enable us to capture children’s comprehensive surroundings (family, neighbourhood, and school).

Third, the EGRA provides a broadly defined measure of children’s preschool experiences, as measured by a single question on preschool attendance. Although the question offers the

[82] This approach contributed to the higher enrolment in O-Class (59%) than the national average (35%), according to the National Education Statistics (MOE, 2015-2016).

example of early learning institutions, including nursery, kindergarten, and O-Class, it is not possible to obtain specific information, such as which institution they attended, type (public or private) and quality (e.g., class size, teacher-child interaction) of the institution, when they started preschool, etc. The possibility of recall problems cannot be excluded, although trends in the preschool enrolment of the EGRA sample are comparable with those from the National Education Statistics.

While I have limited quantitative information on quality, there are a number of elements worth considering for future research. Having information about specific quality indicators allows us to compare the quality of public preschool in Ethiopia to other formal or informal providers, and to analyse the effectiveness by different preschool service provision and the role of quality indicators (e.g., class size, teacher quality, adult-child interactions) in improving students' achievement. For example, research in LMICs has demonstrated the importance of preschool quality and, in particular, the importance of engaged and caring child-teacher interactions in predicting preschool benefits (Aboud, 2006; Araujo et al., 2016). Further research on preschool quality in low-resource settings will be instrumental in extracting relevant policy implications, as it will enable us to identify areas that need further support, change, or regulation, and inform effective resource allocation.

Fourth, the present study is limited to early literacy measures provided by the EGRA as child outcome indicators. Having access to measures assessing other areas of child development, such as early numeracy, socio-emotional development, executive function, and motor development, would have enriched the study findings through the lens of holistic child development.[83] In addition, assessment for early grade reading that is highly sensitive to local languages adds complexity to my analysis and interpretation. Although selected languages are used by more than 90 percent of the population, there is also a limitation in measuring child development among language minority groups. For example, REBs have adapted O-Class materials for the *Saho* and *Kunama* languages in Tigray, for *Argoba* in Afar, and for *Berta, Gumuz,* and *Shinasha* in Benishangul-Gumuz (Rossiter et al., 2018), but these languages have not been used in any national assessment to monitor children's developmental progress.

[83] I attempted to access data on the Early Grade Mathematics Assessment (EGMA) Ethiopia 2014, but there is no permission for public access.

Finally, the current study used strategies for mitigating selection bias attributable to observed sources of preschool attendance, which yields a more robust estimate of the associations between preschool attendance and learning outcomes than a simple comparison between children who attended preschool and those who did not. Given that, in most real-world cases, it is not possible to control for unobserved child and household characteristics that may affect preschool attendance and learning outcomes, the results of the current study cannot be given a direct causal interpretation but does present a rigorous estimate which shows consistent results with a robustness check using propensity score matching. In spite of these limitations, the two datasets leveraged over the reform period make it an important resource for understanding the patterns of association between preschool and child outcomes. None of the studies in Ethiopia and Sub-Saharan Africa deal with this pattern that could better inform policy and scaled-up service provision.

2.11 Conclusion

Many LMICs have recently expanded access to pre-primary education as an instrument for promoting human capital creation and accumulation. Evidence from the U.S. and high-income countries suggests that these investments yield large returns (Berlinski et al., 2008, 2009; Cascio, 2009; Phillips et al., 2017), but evidence from low-income countries is elusive (Bouguen et al., 2015; Brinkman et al., 2017). Evidence is particularly lacking in Sub-Saharan Africa, where the potential for policy interventions in early childhood development is much greater than in any other region.

The present study has contributed to filling this gap in the literature by exploiting a large-scale expansion in access to public preschool in Ethiopia from 2010 to 2016. Leveraging two large, representative EGRA datasets that straddled the early learning reform period, I estimated the association between preschool attendance and students' early grade reading outcomes throughout Grade 2 and Grade 3. I find a positive relation between preschool attendance and improved academic achievement only after the expansion, as measured by test scores on the EGRA sub-tasks (e.g., oral reading fluency), and the proportion of non-readers and proficient readers. These associations varied by regions and languages.

Overall, patterns in 2010 overturned those in 2016, which indicates that the role of preschool was strengthened during the reform period. While it is encouraging that the association has become more pronounced over a period characterised by heightened public interest in ECE, we may need further research on this pattern that focuses on the comparison group—those who still do not benefit from early learning opportunities. The attention of policymakers should be not only on improving the quality of current service provision to maximise the gains from preschool but also on promoting the inclusion of communities that still do not have access to any form of preschool.

This study leaves many questions unanswered about the conditions under which scaled-up preschool can yield meaningful and sustained benefits. While evidence exploiting significant policy shifts with non-experimental designs is instrumental, future research using household-level data and experimental designs may provide more definite answers about the impact of expanding access to pre-primary education on children's learning. This study also captures the inception stage of early learning reform, but the reform in Ethiopia is far from static, and policy efforts are increasingly oriented toward improving quality. Studying the effects of these improvement efforts will inform potential effective ways to stimulate human capital accumulation in many LMICs.

3 CHAPTER 3 – Pathway from Early Childhood Education to Primary Education: Exploring Moderation and Mediation in the Preschool Influence in Ethiopia

3.1 Rationale for the Chapter

Having established the relations between expanded access to preschool and student academic outcomes in the previous chapter, the present chapter aims to extend these findings based on the dimension of equity. Educational equity has been placed at the heart of the international development agenda (SDG 4, United Nations, 2015), and therefore it is critical to distinguish between 'equity' and 'equality'. According to Jacob and Holsinger (2008), equality is defined as 'the state of being equal in terms of quantity, rank, status, value or degree', while equity is considered 'the social justice ramifications of education in relation to the fairness, justness, and impartiality of its distribution at all levels or educational subsectors' (p. 4). Every child grows up in their own unique circumstances and has particular learning needs. If we provide all students with equal learning environments and resources that stress uniformity, it may not be equitable. To make education equitable, we need to offer students from disadvantaged backgrounds or those with special learning needs more intensive interventions and additional schooling. This may not be equal, but it is equitable, as it provides all students the resources and learning opportunities they need to achieve desired outcomes. The key to educational equity is to ensure that all children have a fair chance to develop basic skills and to improve their school readiness, regardless of their socioeconomic, demographic or geographic characteristics.

Equity in education therefore can be conceptualised from different perspectives in a given social and economic context. Any attempt to measure equity cannot be separate from a normative framework of fairness and justice (see UNESCO, 2018, for details). In the current study, I focus on the major sources of educational inequality pointed out in previous work in LMICs (e.g., Lewin, 2009; Banerjee et al., 2008), such as gender, poverty, location, language and ethnicity, which are the key dimensions to be taken into account to achieve equity. This is particularly pertinent to the context of education reform in Ethiopia, where the government used pre-primary education as a means to increase equity at the point of entry to the education system (MoE, 2015). The current chapter delves into whether the relations between preschool and child outcomes differ by the above dimensions and how they changed after the large-scale expansion of O-Class. The previous chapter describes impartial access to preschool based on gender, poverty, location, caregiver's education and language, whereas this chapter assesses the consequences of such impartiality based on learning outcomes. While it is useful to quantify the overall effectiveness of ECE, understanding the specific contextual sources of the differential effects is critical for policymakers in low- and middle-income countries (LMICs), who need to recalibrate their countries' education policies to address the different needs of young children from diverse backgrounds and identify strategies for optimal resource allocation.

3.2 Introduction: Inequalities in ECE

Investment in early childhood education (ECE) is often motivated by its potential to reduce inequalities associated with growing up in disadvantaged circumstances. The risk of poverty-related developmental losses is high; it affects an estimated 250 million children, about half of those under age five in LMICs, and more than two-thirds of the children in Sub-Saharan Africa (Black et al., 2017). With growing evidence that the benefits of ECE are significantly greater for vulnerable and disadvantaged children (Engle et al., 2011; Heckman et al., 2010; Magnuson & Duncan, 2017), governments, multinational organisations, and NGOs are promoting the expansion and improvement of ECE in LMICs as one means of narrowing the learning gaps between children from advantaged and disadvantaged backgrounds (Black et al., 2017; Sayre et al., 2015).

However, empirical evidence on the 'equaliser' role of ECE is largely based on small-scale, high-quality interventions that target low-income families and mostly took place in high-income countries. Little is known about whether a particular group, such as children from the poorest households or those living in rural areas, benefits significantly more from ECE in LMICs. In fact, emerging evidence in these countries presents mixed findings about the benefits of preschool in reducing learning disparities between the rich and the poor. Some studies show larger benefits for disadvantaged children in terms of school readiness (e.g., Brinkman et al., 2017; Jung & Hasan, 2014), while others suggest a null or even negative influence on the academic achievement of children from poor households (e.g., Berg et al., 2013; Bietenbeck et al., 2017; Bouguen et al., 2014).

Ethiopia has been in the process of an extensive ECE policy reform since 2010. The government's primary goal of this investment in ECE is to provide early learning opportunities that promote school readiness and reduce poverty-based inequities at school entry (MoE, 2015). During a six-year period when the gross enrolment rate in pre-primary education rose from 4.8 percent to 50 percent, many previously excluded young children from impoverished backgrounds were included in the country's education system. However, in the absence of rigorous evaluation, it remains unclear whether expanding access to preschool ensures that poor children will be more prepared for formal schooling and reduces the gaps in school readiness between rich and poor children.

The present study provides new evidence on the influence of preschool across a variety of child and family characteristics in Ethiopia. Using data from the Early Grade Reading Assessment (EGRA) 2010 and EGRA 2016, which were conducted during Ethiopia's large-scale preschool expansion, I first examined whether patterns of the relationship between preschool attendance and students' academic achievement differ by gender, urban and rural location, father's or mother's literacy, and home reading resources. The current study focuses in particular on identifying a particular group for whom the benefits of preschool were greater and whether these patterns changed after the expansion of public preschool (O-Class in Ethiopia). Second, the present study adds to a limited but growing literature that explores the sustained benefits of preschool for child outcomes relative to their subsequent schooling experiences. I specifically explored how the relationship between preschool attendance and students' academic achievement is mediated by subsequent schooling environments, as measured by primary school characteristics. Guided by Bronfenbrenner's (1979, 1986; Bronfenbrenner & Morris, 2006) bioecological theory, I explored the multiple sources of the variations in preschool influence from different environments (family, school, community, and policy) and examined these dynamic relations by exploring questions of moderation and mediation in multiple environments.

This paper makes three key contributions. First, it presents the first analysis of the differential influence of preschool on academic achievement using a regionally representative sample in Ethiopia. Although exploration of moderation in ECE research is far from novel, the disadvantaged nature of the sample population (e.g., all children are from low-income families or rural villages) in prior work limits researchers' ability to examine cross-characteristics or cross-site variability in effectiveness (Barnett & Belfield, 2006; Burger, 2010; McCoy, Morris, et al., 2016).[84] Second, focusing on the large-scale expansion of preschool in Ethiopia shows

[84] In the previous studies conducted in Ethiopia (Woldehanna & Araya, 2012, 2017), the sample population was children from urban areas, since only children from affluent backgrounds were able to access (private) preschool.

the trends of the differential benefits of preschool by sub-group before and after the reform. The reform accompanied a major shift in Ethiopia's ECE landscape from an elite to a mass system, which is a context not studied before in low-resource settings. Third, given the importance of sustaining environments between pre-primary and primary education (Bailey et al., 2017), the study offers the first exploratory analysis of the role subsequent schooling environments play in mediating the relationship between preschool attendance and student outcomes in the LMIC context. The rest of this paper is structured as follows: Section 2 summarises the relevant literature; Section 3 outlines the purpose and research questions of the present study; Section 4 presents the key variables used in the analysis; Section 5 provides the empirical approach; Section 6 presents the results from the analysis; Sections 7 and 8 discuss the findings and conclusions.

3.3 Relevant Literature

Educational inequalities start even before children begin school, driven primarily by disparities in wealth. An extensive body of research from developed (Brooks-Gunn & Duncan, 1997; Currie, 2009; Feinstein, 2003) and developing countries (Grantham-McGregor et al., 2007; Naudeau et al., 2011; Rolleston, James, & Aurino, 2013; Schady et al., 2015; Walker et al., 2011) suggests that there are steep socioeconomic status (SES) gradients in early childhood development, as most of the observable cognitive gap between wealthier and poorer children emerges prior to any formal schooling and often widens with age. Early cognitive and non-cognitive skills are, in turn, important determinants of success in terms of subsequent educational attainment (Walker et al., 2011; Currie, 2009), adult health (Campbell et al., 2014), criminality (Currie, 2001), and the probability of employment and level of future earnings (Chetty et al., 2011; Gertler et al., 2014; Heckman, Malofeeva, et al., 2010).[85] The evidence highlights the fact that disadvantages found at an early age will result in the intergenerational transmission of poverty and income inequality (Barnett & Belfield., 2006; Heckman & Mosso, 2014) if not addressed by some measure before children reach primary school age.

Theories and evidence from the U.S. argue that policies targeting early childhood development may be cost-effective solutions for promoting school readiness and reducing income-related

[85] The listed evidence on long-term effects come from the U.S., except evidence from Jamaica by Getler et al. (2014).

inequalities at school entry.[86] Compared to parental care at home, preschools are thought to prepare children more fully for a structured primary school environment, particularly disadvantaged children whose low-educated parents might not be able to provide similar stimulation at home (Brooks-gunn, 2003; Duncan & Magnuson, 2013). Attending an ECE programme with a rich learning environment may partially compensate or substitute for lower levels of parental investment and less stimulating learning conditions (Ramey & Ramey, 1998). Various early interventions that target disadvantaged children, such as an intensive preschool programme (Campbell et al., 2002, for the Abecedarian Project; Schweinhart et al., 2005, for the Perry Preschool Study) or a large-scale comprehensive early intervention (Deming, 2009; Ludwid & Miller, 2007, for Head Start), establish the long-term effects early childhood intervention has on adulthood, even after an initial period of fadeout. Bolstered by the empirical evidence, along with dynamic skill formation models (Cunha & Heckman, 2007), the rationale for investment in ECE as the most effective tool for reducing gaps in child development has grown stronger. However, compared to the evidence from the U.S. and high-income countries, relatively little evidence has systemically documented whether ECE programmes in LMICs are able to compensate for the socioeconomic gradients in cognitive development, and hence to address one major cause of educational inequalities.

Prior studies point to a range of family and individual characteristic that may lead to differential effectiveness of ECE in rich countries, including family income, race/ethnicity, maternal education, and home language (Heckman, 2006; Magnuson & Duncan, 2006, 2014; Halle et al., 2009; Duncan & Magnuson, 2011; Reardon & Portilla, 2016). Similarly, in poorer countries, there is a variety of causes of learning disparities, including poverty, gender, geographic location, disability, malnutrition, and ethnic and linguistic minority status, which often interact with one other to reinforce disadvantage (Altinok, 2013; Burger, 2011; Moloi & Chetty, 2010; Rolleston et al., 2013).

In Ethiopia in particular, multiple sources of inequality remain for children from the poorest households and/or rural areas, girls (in secondary school), and children whose parents have little formal education (Rolleston et al., 2013; Pankhurst et al., 2018; Woldehanna et al., 2017). A study based on the Young Lives longitudinal data that compared four countries—Ethiopia, Peru, India (Andhra Pradesh), and Vietnam—demonstrated that students from the richest

[86] School readiness is the degree to which a child is prepared to learn and succeed in school (Ackerman & Barnett, 2005).

quartile made more progress than those from the poorest quartile in mathematics during their transition from home (age 5) to primary school (age 8) (Rolleston et al., 2014).[87] By age eight, there is a gap of 33 percentage points between richer (87%) and poorer (54%) children in Ethiopia in their ability able to answer 'how much is 2 multiplied by 4'. Moreover, urban 8-year-olds are over five times more likely to be able to read sentences than rural 8-year-olds (Rolleston et al., 2013). In addition to the poverty-based learning gap, Singh (2014) found differential productivity of a single-year of schooling in the four countries that largely accounts for the cross-country divergence in learning at age 8, which was evident at age 5 and grew substantially in the first 2-3 years of schooling.[88]

3.3.1 Empirical Evidence on the Differential Effects of ECE

Between girls and boys. Despite the prevailing norms of gender equity in education, girls and boys continue to perform differently in school. Evidence from the U.S. suggests that gender disparities in academic achievement may mirror differences in early development and learning opportunities (DiPrete & Jennings, 2012; Entwisle et al., 2007). However, evidence is mixed regarding the extent to which ECE may mitigate gender differences across different ages. According to a meta-analysis of 23 ECE programmes in the U.S., the effects of ECE programmes on cognitive outcomes are generally similar for girls and boys, with greater benefits for boys on school outcomes such as grade retention and special education placement (Magnuson et al., 2016). Prior work evaluating individual programmes reported larger gains from ECE for boys in the early grades (Deming, 2009; Muschkin et al, 2018) and for girls in adulthood, while other work reported no gender differences (Weiland & Yoshikawa, 2013).

The differential benefits by gender were often mixed within the same intervention. In evaluating the Carolina Abecedarian Project, García, Heckman, and Ziff (2018) found that, across the positive impacts of ECE during the life cycle for both genders, boys experienced greater benefits on long-term outcomes such as the labour market, employment, health, and reduced participation in crime; girls experienced greater benefits on short-term outcomes such as cognition, achievement, and educational attainment. Although there is relatively limited evidence on gender differences outside the U.S., most prior studies consistently report that

[87] Young Lives is an international study of childhood poverty that followed the lives of 12,000 children in four countries (Ethiopia, India, Peru, and Vietnam) over 15 years.
[88] The productivity of a school year is measured by 'learning gains per grade completed' in value-added models (Singh, 2015).

there are no significant gender differences in preschool benefits in Argentina (Berlinski et al., 2009), rural Guatemala (Bastos et al., 2017), Kenya and Tanzania (Biethenbek et al., 2017), Turkey (Agirdag et al., 2015), and Uruguay (Berlinski et al., 2008). In Ethiopia, using the urban sample of children from the Young Lives Study, Woldehanna and Gebremedhin (2012) found that preschool attendance had a slightly greater influence on the cognitive development of girls than of boys when the children were age 5, but that preschool benefits were slightly greater for boys than girls when the children were age 8. Eight-year-old boys who attended preschool also had a higher chance of enrolling in primary school and to make a timely grade progression than 8-year-old girls.

Between advantaged and disadvantaged. Although earlier studies in the U.S. emphasised the role of ECE in reducing learning disparities between rich and poor children, most of them evaluated the effectiveness of early interventions that targeted low-income populations; that is, on average across the studies, about 90 percent of the samples were from families in poverty (Leak et al., 2010).[89] In recent years, a growing number of studies have been conducted using national or state-wide sample data (e.g., Early Childhood Longitudinal Study). Researchers have attempted to identify whether there are particular groups for whom the benefits of preschool are larger and more persistent among the representative sample. In line with predictions from compensatory models, these studies found that the preschool benefits are more pronounced among low-income children at school entry (Weiland & Yoshigawa, 2013) and through adolescence (Casio & Schanzenbach, 2013). For other individual-level characteristics, evidence is consistent of greater preschool benefits for racial/ethnic minority children (Weiland & Yoshigawa, 2013; Gormley et al., 2008), dual language learners (Bloom & Weiland, 2015; Puma et al., 2010a), and rural communities (Fitzpatrick, 2008). In a review of ECE studies, largely from high-income countries, Burger (2009) found that preschool programmes might compensate for social inequalities across highly diverse settings. Longitudinal studies conducted in Germany (Socio-Economic Panel), the U.K. (National Child Development Study, British Cohort Study), and the U.S. (North Carolina More at Four Pre-K Program) document that children from disadvantaged family backgrounds, as defined by poverty level, immigrant status, special needs, English proficiency, and chronic health

[89] The Head Start and other targeted programmes (Perry Preschool and the Abecedarian Project) have extensively studied the heterogeneity in effects, but selection bias is a major problem, given the disadvantaged nature of the sample population (Barnett & Belfield, 2006). This limits researchers' ability to examine cross-site or cross-characteristics variability in effectiveness (McCoy et al., 2017).

conditions, made more progress than their more advantaged peers when they attended preschool (Burger, 2009).

Empirical evidence from LMICs is more mixed on the following three patterns: (1) disadvantaged children benefit more from preschool than their advantaged peers; (2) children from different family backgrounds benefit equally; and (3) disadvantaged children benefit less than their advantaged peers. First, as evidence from high-income countries claims, several studies found that preschool had greater benefits for children from disadvantaged families, as defined by SES, parental education level, parenting practices, and urban and rural location. In evaluating the scale-up preschool initiatives in Argentina and Uruguay, disadvantaged children obtained larger benefits from the expansion of public preschool. The positive effects of preschool attendance were more pronounced among children from low-SES families on their third-grade test scores in Argentina (Berlinski, Galiani, & Getler, 2009), and among children with low-educated mothers on their level of educational attainment at age 15 in Uruguay (Berlinski, Galiani, & Manacorda, 2008).

A study looking at the effects of a community-based ECE intervention in rural Indonesia found that the gap between rich and poor in the treated villages decreased on multiple domains in child development, including cognitive and non-cognitive skills. By contrast, in non-treated villages, this rich-poor gap either increased or stayed constant (Jung & Hasan, 2014). In the follow-up evaluation of the same intervention, Brinkman et al. (2017) found that, for children from more disadvantaged backgrounds—poorer households or poorer parenting practices—the preschool benefits were larger and did persist over time for children's language and cognitive development and social competence. In terms of outcomes in educational attainment, a study of the ECE programme in rural India found a particularly strong and positive relationship between preschool and primary school participation among children from households below the poverty line (Hazarika & Viren, 2013).

Second, prior studies found no differential effects of preschool by student background. In the Sub-Saharan Africa context, a study using nationally representative data from Kenya and Tanzania found no differential gains from preschool on students' educational attainment and cognitive development between richer and poorer households, rural versus urban location, and mother's education (Bietenbeck et al., 2017). Similarly, work by Bastos et al. (2017) that evaluated preschool construction in rural villages in Guatemala found no consistent differences in preschool benefits on educational attainment between advantaged and disadvantaged communities, as defined by the share of adults with no education, an indigenous population, and the prevalence of chronic malnutrition. Rather, the positive benefits of preschool attendance were lower in communities where a greater share of adults had no education (Bastos et al., 2017).

Third, some evidence raises concerns that, rather than ameliorating inequalities, preschool can further extend the privileges of more advantaged children or the more deleterious effects for less advantaged children. In South Africa, a study assessing the influence of a one-year pre-primary class (R-Class) found that preschool had virtually no measurable benefits in terms of academic performance in schools from the lower wealth quintile, while there were more discernible benefits for those from the highest wealth quintile (Berg et al., 2013). Similarly, a study looking at Turkish students' performance in the Programme for International Student Assessment 2012 reported that, although preschool was related to higher academic achievement for all participating students, students from wealthy families obtained greater benefits than students from poorer families (Agirdag et al., 2015). Strikingly, in a randomised experimental study of a large-scale ECE programme in Cambodia, Bouguen et al. (2013) found a negative effect of a new ECE programme on the cognitive development of children; the largest negative effects were found among children from poorer household and those with less educated parents.

In all, despite the potential of ECE to close existing equity gaps, the best ways to deliver greater benefits for marginalised children remain elusive. It is unclear what causes this variation within and across countries. Researchers point out that low-quality preschool, especially in poorer communities—for example, with a weak infrastructure and unqualified or untrained pre-primary teachers—may determine whether attending preschool delivers beneficial effects across a variety of family characteristics (Barnett, 2008; Berg et al., 2013; Burger, 2010). Others surmise that preschool is too late to intervene, in that much of the brain develops before age three, the time when many children are entering preschool (Richter et al., 2017). Thus, the accumulated developmental losses from birth are unlikely to be compensated by preschool interventions (Burger, 2010).

3.3.2 Empirical Evidence on the Variation in Effects of ECE by Subsequent School Environments

A growing body of ECE research focuses not only on early gains from the preschool experience, but also on how well the benefits of preschool are sustained during subsequent schooling experiences.[90] Earlier studies in the U.S. indicated that the benefits of preschool fadeout more quickly among disadvantaged groups such as black children because they are more likely to attend poorer quality primary schools (Currie & Thomas, 2000; Lee & Loeb, 1995). Given that the benefits of preschool depend in part on the quality of the primary school, some argue that the benefits of a preschool experience would fadeout unless reinforced by subsequent good school experiences (Bogard & Takanishi, 2005). This aligns with the cumulative model of human capital acquisition (Cunha & Heckman, 2007), also known as 'dynamic complementarity', which suggests that the benefits of preschool may be enhanced when followed up by a high-quality school environment. Recent work focuses on the importance of 'sustaining environments', or the inputs and features of primary school experiences, as critical pathways to preserving children's preschool benefits (Bailey et al., 2017; Phillips et al., 2017).

To examine whether subsequent schooling environments contribute to sustaining the effects of preschool, a wide range of school characteristics has been investigated, mainly in the U.S. At least three patterns emerged from the existing literature on how to define, measure, and treat subsequent schooling experiences relative to the link between preschool and students' outcomes: (1) either school-level or class-level factors; (2) either structural or process school quality indicators; and (3) treating these factors as either 'moderator(s)' or 'mediator(s)'.

First, the distinction between how classroom-level and school-level pathways affect student achievement is based on the proximity to the instructional interactions (Curenton, Dong, & Shen, 2015). For instance, classroom-level pathways focus on the proximal factors, including student-to-student (peer) learning interactions, class size, and teachers' instructional approach, such as spending more time on advanced content rather than on basic math and literacy content (Claessens et al., 2014). By comparison, school-level pathways draw on the distal factors that affect the teaching and learning process. For example, school-level pathways include schoolwide infrastructure and resource level (e.g., school assets, expenditure per pupil), indicators of average schoolwide student achievement (e.g., proportion of students at or above

[90] Researchers explored preschool effects that varied by the 'characteristics of preschool or ECE programme', such as the duration and intensity (Loeb et al., 2005; McCoy et al., 2016) and teacher-child interactions in the ECE programme (Araujo, Carneiro, Cruz-Aguayo, & Schady, 2017; Mashburn et al., 2008; Hu et al., 2017). However, due to the absence of data on preschool characteristics or quality, the current study was not able to explore this aspect.

academic proficiency), and indicators of mean student SES (e.g., proportion of students who received free/reduced-price lunch) (Hanushek, 2003).

Second, the school- and classroom-level pathways are further diversified by the structural and process aspects of school quality. Although there is no consensus on what constitutes school quality (Glewwe, Hanushek, Humpage, & Ravina, 2013; Sammons, 2009), it often is measured by structural and process characteristics that are thought to stimulate student learning, especially in early grade settings (Howes et al., 2008; Layzer & Goodson, 2006; Sylva et al., 2006; Thomason & La Paro, 2009).[91] Structural quality refers to the overarching structures needed to ensure quality in school, including the physical environment (school buildings, outdoor space, learning materials), teachers' qualifications and training level, pupil-to-teacher ratios, and standards that regulate the learning environment and workforce conditions (Bryant, Zaslow, & Burchinal, 2010; Philips et al., 2000; Pianta et al., 2005). Process quality consists of student's learning experience through their interactions with teachers, peers, and the learning materials they are engaged with (Pianta et al., 2005; Moss & Dahlberg, 2008). The key dimension of process quality includes teachers' instructional support (e.g., quality of feedback), emotional support (e.g., positive climate), and classroom organisation (e.g., instructional learning formats) (Pianta, Laparo et al., 2008).

Third, the role of subsequent schooling environments can be viewed as either a 'moderator' or a 'mediator' (see Baron & Kenny, 1986), depending on the researcher's hypothesis. To illustrate, a moderator is one variable that affects the direction and/or strength of the relation between preschool experience and later learning outcomes, while a mediator is one variable that explains how and why such a relation was established (or failed to be). If subsequent school environments moderate this relation, it implies that the way the child with preschool experience is engaged in primary schools is different from the engagement of a child without preschool experience, which changes the nature of the link between preschool and student outcomes. By comparison, if subsequent school environments serve as mediators, they are part of the intervening mechanism within the established relation. The effect of preschool experience is thus transmitted via the children's engagement in primary school, which may provide a conducive learning environment, to later outcomes.

[91] Note that the literature on structure and process quality is largely from studies in ECE, including pre-kindergarten, kindergarten (U.S.), or reception class (U.K.).

Considering all these variations in conceptualizing and operationalizing the school characteristics, it may not be surprising that the empirical evidence on whether subsequent school experiences play a complementary or compensatory role is mixed. The first strand of research focused on assessing ‘school-level’ factors related to child outcomes. Focusing on the impact of Head Start, Currie and Thomas (2000) found that the Head Start programme’s fade-out effects among black students depended on the fact that these students were more likely to attend poor-quality schools where students’ average test scores were low. Relatedly, Curenton et al. (2015) found that the association between ECE attendance and fifth-grade academic achievement was partially mediated by aggregate schoolwide achievement. Using the Chicago Longitudinal Study, Reynolds, Ou, & Topitzes (2004) found that attending a high-quality primary school (e.g., a magnet school) mediated the long-term benefits of preschool on educational attainment at age 20.[92] Among the multiple indirect paths of students, family, and school characteristics (mediators), school factors accounted for about one-third of the total indirect effect (Reynolds et al., 2004). Most recent research on Head Start revealed that its benefits were more pronounced when followed by access to primary schools with higher funding levels, with particularly larger benefits among children from the poorest families (Johnson & Jackson, 2018). Aligned with a dynamic complementarity model (Cunha & Heckman, 2007), these findings suggest that early investments that are followed by sustained educational investment over time can effectively reduce inequality between the rich and the poor (Johnson & Jackson, 2018).

The second strand of research focused on evaluating ‘classroom-level’ factors that may play a role in the relationship between preschool and learning outcomes. Using the U.S. Early Childhood Longitudinal Study, Magnuson et al. (2007) showed that the persistence of preschool benefits was tied to both class size and quality of instruction; unexpectedly, preschool gains persisted for children who experienced *larger* class sizes and *lower* levels of reading instruction but were eliminated for children who subsequently experienced *smaller* class sizes and *higher* levels of reading instruction. By comparison, Bassok et al. (2018) found that primary school characteristics, measured by length of school day, class size, transition practices, and exposure to advanced content, did not alter the relationship between preschool and third-grade cognitive and behavioural outcomes. Similarly, Jenkins et al. (2018) found that

[92] Magnet schools are selective elementary schools (private) that provide specialised programs across the curriculum. A higher proportion of students in these schools exceed national norms in reading and math achievement (Hickey & Reynolds, 2002).

advanced content and high-quality instruction in the primary grades did not moderate the fadeout of preschool effects, except for some mitigation induced by targeted teacher professional supports.

Overall, a large body of the literature exploring school- or classroom-level pathways measures school environments by the structural quality indicators of school, which are easier to measure than the process quality indicators. With respect to the moderation *versus* mediation role of subsequent school experience, except for two studies using the mediation model (Reynolds et al., 2004; Curenton et al., 2015), most studies used the moderation model to see whether primary school characteristics altered the extent to which preschool affects students' outcomes. Meanwhile, inconsistency in the findings across the studies, which often reported no moderation effect, raised some methodological issues for such research, such as the difficulty in measuring school or classroom quality and the non-random assignment of post-preschool environments (Bailey et al., 2017).

It should be noted that all evidence presented here is from the U.S. and, to my knowledge, no study has explored the subsequent schooling experience in the LMICs.[93] Two studies on ECE in Ethiopia applied the moderation or mediation model, but these studies did not account for post-preschool experience in primary school. Both used the Young Lives Study, which captured the period when only urban children had access to private preschool. A study by Woldehanna and Araya (2017) examined the differential influences of preschool by preschool characteristics, including duration (one to three years) and preschool type (private, community, and governmental), on educational attainment at age 18. The results indicated that the contribution of preschool was more pronounced for children who attended preschool for at least three years or for those who attended private preschool. Another study by Woldehanna (2016) looked at the indirect effects of preschool on the relation between family background and students' cognitive outcomes. The author found that preschool attendance partially mediated the influence of family background—household wealth, parental education, and regional location—on students' cognitive achievement. To fill these gaps in the knowledge, the present study employed a mediation model to test the hypothesis that preschool attendees' early grade achievement would be mediated by subsequent schooling environments. Given that

[93] A relevant example was found in China, but the study used only structural and process quality within preschool. (B. Y. Hu, Zhou, Chen, Fan, & Winsler, 2017) examined the relationship between financial resources in ECE and student academic outcomes, mediated by teacher-child interactions in the pre-primary classrooms.

skill development is a multi-stage process in which investments at previous stages interact with and complement the current one, it is reasonable to investigate whether the relation between preschool and students' early grade performance relates to subsequent primary school experiences.

3.4 The Present Study

Building on the positive relations between expanded access to preschool and children's literacy outcomes established in Chapter 2, the present study aims to examine whether the preschool benefits vary on two theoretically motivated dimensions: (1) the socio-demographic characteristics of child and family; and (2) subsequent schooling environments. Using the Early Grade Reading Assessment (EGRA) administered in 2010 and 2016 to a regionally representative sample of Ethiopian students, the following questions will be addressed:

1. Do some child sub-groups, as defined by child gender, urbanity, paternal and maternal literacy, and home reading resources, benefit significantly more from preschool attendance than others, before and after the early learning reform?
2. How are the relationships between preschool attendance and second- and third-grade reading achievement mediated by subsequent schooling environments?

I first aim to identify the particular group for whom the benefits of preschool are greatest (as measured by the probability of being a non-reader or a proficient reader) and whether these patterns differ before and after the early learning reform. The definition of sub-groups—child gender, urbanity, paternal and maternal literacy, and home reading resources, which are available in the dataset—is guided by earlier studies in the region (McEwan, 2014; Bashir et al., 2018) and in Ethiopia (Piper, 2010; Rolleston et al., 2013; Woldehanna, 2016). Drawing on previous evidence, I hypothesise that the benefits of attending preschool will be significantly greater for children from disadvantaged backgrounds: being a girl, living in a rural area, living with an illiterate father and/or mother, and not having any reading materials at home. Second, I explore how the relationships between preschool attendance and students' early grade reading performance (as measured by EGRA test scores) are mediated by their primary school schooling environments. My hypothesis is that school characteristics (e.g., class size, the availability of textbooks, and school principal's leadership) play a role in sustaining the benefits of preschool along the pathway between preschool attendance and students' learning outcomes. Due to a lack of previous evidence and limited measures of school quality, my examination of the mediating role of primary school environments is considered exploratory.

The present study contributes to the existing literature in a number of ways. First, to my knowledge, this is one of the first studies to use a regionally representative sample to assess the degree to which the relations between ECE attendance and early grade reading skills differ, based on the socio-demographic characteristics of students in Ethiopia. Prior studies on ECE in Ethiopia often were limited to an urban sample (Woldehanna & Araya, 2017) or to one targeted region or community (Dowd et al., 2016), thus certain restrictions exist in conducting sub-group analysis. Second, while the existing literature focuses on the direct effects of preschool or on whether subsequent schooling experiences alter the magnitude of these effects, the present study is the first to examine the mediating pathways of students' subsequent school environments and learning outcomes, especially in the LMIC context. The present study provides a better understanding of targeted approaches to preschool expansion aimed at ensuring equitable access and learning for the disadvantaged, and of strategies for aligning the pre-primary and primary school experiences in ways that sustain the benefits of early learning.

3.5 Data and Variables

The present study used the dataset from the Early Grade Reading Assessment that was constructed in Chapter 2. Sample, key explanatory variable, outcomes variables, and control variables used for the current analysis are same as in the previous chapter (see Section 2.6). To address my second research question applying the mediation analysis, I added measures of school environments from the same EGRA dataset as follows:

Measures for school environments. The hypothesised mediating pathway of subsequent school environments is intended to be operationalised using variables that reflect both the structural and process quality provided for children in the school. To measure school characteristics, I used information collected by the EGRA questionnaire on school principals, focusing in particular on activities that support mother tongue instruction (AIR, 2016): principals' qualifications and experience (having a bachelor's degree or higher, having training in early grade reading); principals' involvement in reading instruction (supporting teachers in reading instruction, reviewing teachers' lesson plans every week, satisfying students' early grade reading achievement, managing classroom observations, conducting oral examinations to monitor early grade reading progress); and the schools' resources and staff management (having a school library, having new mother tongue textbooks, and having teachers trained in early grade reading).

Some measures of school environment are relevant to the principal's support for teaching and learning practices, which may lead to schools' improved process quality (e.g., staff-child interaction, school leadership, school climate). In the broader literature on school effectiveness, these practices may relate to the characteristics of an effective school, such as a *participative approach to leadership* (Sammons et al., 1995), and to key elements in school improvement, such as *self-evaluation* (MacBeath, 2010), as they support the identification of barriers to teaching and learning. The process quality indicators are also connected to schools' structural quality (e.g., class size, learning materials, teachers' qualifications). In his study on school leadership in Ethiopia, (Abebe, 2012) stressed that school principals play a pivotal role in school management, including the utilisation of textbooks and teaching materials, teacher management, and monitoring support for children in disadvantaged circumstances.

Other measures of school environment are relevant to structural quality of school such as principals' qualifications, the availability of textbooks, and school library. Across a substantial body of literature, evidence on the effectiveness of textbook and principals' qualification on students' outcome was inconclusive (Conn, 2014), while some studies indicated that a school library has statistically significant effects on the increased time the students spend in school (Glewwe et al., 2013). Meanwhile, evidence on the effectiveness of structural school quality is limited but growing in Ethiopia; for example, students' reading fluency is positively associated with the provision of textbooks (Woldehanna, Jones, & Bekele, 2005; Piper, 2010), and students' early math achievement is positively associated with principals' qualifications and the availability of textbooks (NEAEA, 2014).

3.6 Empirical Strategy

3.6.1 Research Question 1: Interactions

To estimate how relations between preschool attendance and learning outcomes differed by child and family characteristics, I estimated a series of models in which I allowed for an interaction between preschool and one potential moderator. I started by recalling the multivariate logistic regression model used in the previous chapter:

$$y_{is} = \beta_0 + \beta_1 PRE_{is} + \beta_2 X_{is} + \beta_3 F_{is} + \epsilon_{is} \qquad (1)$$

where Y_i represents the early grade reading achievement for a student i in school s; PRE_{is}, represents a binary variable of preschool attendance; X_{is} and F_{is} represent a set of control variables, each denoting student- and family-level characteristics, including regional dummies; and ϵ_{is} indicates an error term (residual). With respect to my research interest in the differentials by sub-groups, I extended the equation (1) with the *interaction* terms, as expressed below:

$$y_{is} = \beta_0 + \beta_1 PRE_{is} + \beta_2 Female_{is} + \beta_3 PRE_{is} * Female_{is} + \beta_4 X_{is} + \beta_5 F_{is} + \epsilon_{is} \quad (2)$$

$$y_{is} = \beta_0 + \beta_1 PRE_{is} + \beta_2 Rural_{is} + \beta_3 PRE_{is} * Rural_{is} + \beta_4 X_{is} + \beta_5 F_{is} + \epsilon_{is} \quad (3)$$

$$y_{is} = \beta_0 + \beta_1 PRE_{is} + \beta_2 F\_Literacy_{is} + \beta_3 PRE_{is} * F\_Literacy_{is} + \beta_4 X_{is} + \beta_5 F_{is} + \epsilon_{is} \quad (4)$$

$$y_{is} = \beta_0 + \beta_1 PRE_{is} + \beta_2 M\_Literacy_{is} + \beta_3 PRE_{is} * M\_Literacy_{is} + \beta_4 X_{is} + \beta_5 F_{is} + \epsilon_{is} \quad (5)$$

$$y_{is} = \beta_0 + \beta_1 PRE_{is} + \beta_2 BookatHome_{is} + \beta_3 PRE_{is} * BookatHome_{is} + \beta_4 X_{is} + \beta_5 F_{is} + \epsilon_{is} \quad (6)$$

The dependent variable y_{is} is the log-odds of being a non-reader or a proficient reader, where the probability p is given as $p_{is} = \frac{e_{is}^{y}}{1+e_{is}^{y}}$.[94] From equations (2) to (6), the logit regression model includes one additional term that captures the interaction between preschool and gender, urban and rural residence, father's literacy, mother's literacy, or reading materials (books) at home. Each model estimates the simultaneous association of the interaction term between two groups, as opposed to estimating the model separately for each of these groups. Inclusion of the interaction terms thus provides better understanding of the relationship among the study variables, especially whether their joint associations are significantly greater (or significantly less) than the sum of the parts.

Specifically, in the model with interaction, the role of preschool attendance is conditioned on gender, urban-rural residency, paternal or maternal literacy, and home reading resources. This model provides an estimate of the benefits of preschool for each group, such as the differentials between boys *versus* girls, children living in rural *versus* urban areas, children with an illiterate *versus* literate father or mother, and children with fewer *versus* more home reading materials. This estimate will show, for each group, whether there was a gain in the reduced probability of being a non-reader (or increased probability of being a proficient reader when the model used that outcome variable). If these estimates are the same between two groups, it indicates that all

[94] This chapter deals with the outcome variables on the probability of being a non-reader or a proficient reader. The model with student's outcomes measured by EGRA test scores is available upon request.

groups of children benefited equally from preschool attendance. On the other hand, if these estimates are different between the groups, it indicates that a particular group of children benefited more from preschool than others. For instance, the interaction terms allow us to test whether children living in rural areas benefited significantly more than those living in urban areas in terms of early literacy outcomes.

As argued by several researchers (Brambor, Clark, & Golder, 2006; Kam & Frazese, 2005), all constitutive terms should be included when constructing an interactive model specification, otherwise the risk of incurring inferential errors significantly increases. If one of constitutive terms is dropped in the model, all coefficients on the remaining terms are altered either in size or significance, which causes the other parameters of interest ($\beta_1$to β_5) be estimated with bias. The bias is due to the fact that the excluded terms are correlated with those remaining in the model and with the outcome variables. Moreover, the exclusion of those terms implicitly assumes that the intercept of y_{ij} on Pre (β_0) is not conditional on regressors (e.g., being female or father's literacy), thereby imposing a fixed intercept on the model. In turn, when the model specifications include full constitutive terms, researchers should not interpret this as if they are unconditional marginal effects (Brambor et al., 2006).

Importantly, in the models with the interactive terms, the estimated parameters cannot be directly interpreted as 'effects' (Kam & Franzese, 2007, pp. 19-20). In other words, the coefficients β_1-β_3 in equations (2) to (6) cannot tell us the estimated effects on the outcome y_{ij} if they are taken in isolation. Due to interdependencies, each variable involved in interaction terms has multiple effects, depending on the levels of the other variables with which it interacts.[95] For this reason, when calculating the conditional marginal effects of preschool attendance as a function of gender, urban or rural residence, father's literacy, and reading materials at home, I use the first derivatives of equations (2) to (6) with respect to preschool attendance PRE_i (Brambor, Clark, & Golder 2006; Kam & Franzese, 2005) as follows:

$$\frac{\partial y_{is}}{\partial PRE_i} = \beta_1 + \beta_3 Female_i + \epsilon_{is} \qquad (7)$$

$$\frac{\partial y_{is}}{\partial PRE_i} = \beta_1 + \beta_3 Rural_i + \epsilon_{is} \qquad (8)$$

[95] When linked variables are specified to the model including interaction (e.g., pre ## female), Stata command *margins* is taking care of the marginal or discrete changes correctly (Long & Freese, 2014). Since it is not possible to estimate a separate effect for the interaction, researchers will have the marginal effects of the component terms (Williams, 2012).

$$\frac{\partial y_{is}}{\partial PRE_i} = \beta_1 + \beta_3 F\_Literacy_i + \epsilon_{is} \qquad (9)$$

$$\frac{\partial y_{is}}{\partial PRE_i} = \beta_1 + \beta_3 M\_Literacy_i + \epsilon_{is} \qquad (10)$$

$$\frac{\partial y_{is}}{\partial PRE_i} = \beta_1 + \beta_3 BookatHome_i + \epsilon_{is} \qquad (11)$$

Equations (7) to (11) show how the effect of preschool attendance (PRE_i) is a function of β_1, and other terms interacted with PRE_i. For example, in the relationship between preschool attendance and father's literacy, each category of children can be defined as:

$$\frac{\partial y_{is}}{\partial PRE_i.Father\ Illiterate} = \beta_1 + \epsilon_{is} \qquad (12)$$

$$\frac{\partial y_{is}}{\partial PRE_i.Father\ Literate} = \beta_1 + \beta_3 F\_literacy_i + \epsilon_{is} \qquad (13)$$

For children with an *illiterate* father (eq. 12) the effect of preschool is given by β_1, since for those children, $F\_literacy_i$ would be zero, while for children with *literate* father (eq. 13), the effect is the sum of $\beta_1 + \beta_3$. Because β_1 is correlated with β_3, the variance of their sum is $var\ (\beta_1 + \beta_3) = var\ (\beta_1) + var\ (\beta_3) + 2 \times Covar\ (\beta_1.\beta_3)$.[96] Again, our interest is in knowing whether the effect of a child attending preschool is the same for a child whose father is literate and for a child whose father is illiterate, which is expressed below:

$$H_0: \frac{\Delta \Pr\ (y=1\ |x, F_{\_literacy}=0)}{\Delta \text{ Pre}} = \frac{\Delta \Pr\ (y=1\ |x, F_{\_literacy}=1)}{\Delta \text{ Pre}} \qquad (14)$$

To test this hypothesis, I computed the average marginal effects of the association between preschool and student outcomes, averaging only those cases where father's literacy was 0 (as observed for other values) and compared it with the average marginal effects for those cases where father's literacy was 1 (eq. 14). Once I computed the average marginal effects, which were equivalent to a discrete change by each case, I tested whether the effects were equal or significantly different between two sub-groups by estimating the contrasts of margins.[97] The

[96] As argued by Kam and Franzese (2007, pp. 128-129) the heteroskedasticity (non-constant variance) can be easily addressed in the regression by using White's procedure to generate a consistent variance-covariance matrix and, therefore, to correct the coefficient-estimates' standard errors in the estimated model. This is done in Stata using the option robust or vce(robust) for the regression command.

[97] I used the contrast of margin with, for example, difference in preschool and difference in gender: using Stata command *margins r.pre##r.female (or margins pre##female, contrast(nowald pveffects) vsquish),* one can test whether the groups' difference is significant or not.

estimation of average marginal effects with predicted values, especially in the case of nonlinear models with interaction terms, offers a useful interpretative aid for information on the substantial and practical significance (Cameron & Trivedi, 2010; Williams, 2012).

Additionally, I introduced the school fixed effects approach to the multivariate logistic model in order to compute within-school difference by controlling for the potential confounding effects of all unobserved, time-invariant school variables. An analogous procedure of the fixed effects used in the linear model was applied to the logit model with fixed effects; however, instead of maximum likelihood, 'conditional' maximum likelihood should be used to avoid bias caused by the incidental parameters problem (e.g., creating multiple records or dummy variables for schools) (Allison, 2009). Also, the interpretation of the school fixed effects logit model was limited to 'odds ratio'. Estimating marginal effects can be problematic after the fixed effect model is applied because, by default, margins provide information on the probability of a positive outcome assuming that the fixed effect is 'zero', which may be an unreasonable assumption when computing the predictive values (Pforr, 2013; Silva & Kemp, 2016; Williams, 2018).

3.6.2 Research Question 2: Structural Equation Modeling

To address my second research question, I used mediation analysis to explore how a 'mediator' variable (subsequent schooling experience in this study) might influence the relationship between preschool and students' academic achievement. Structural equation modeling (SEM) (Baron, & Kenny, 1986) is an appropriate method for this analysis because of its suitability to unpacking the intervening mechanism based on the relations established in the previous study (see Chapter 2), to testing multiple processes simultaneously (including the direct and indirect paths of all predictors while taking into account a variety of covariates), and to comparing alternative models by assessing model fit statistics. The mediation analysis should be distinguished from the analysis of the moderation effect used in my first research question, which provides insights into whether ECE programmes may operate differently for different sub-groups or in different primary school settings.[98] Analysis of the mediation effect allows me to identify the pathways to achieving a positive relationship (or the obstacles if the intervention failed) through multiple channels in subsequent school environments.

[98] Moderation analysis was conducted to explore whether the association differed by subsequent schooling experiences. However, there was no statistically significant effect of interaction terms (e.g., preschool ## school having continuous monitoring on child literacy skills) on students' outcomes.

Specifically, I used a multiple mediator, single-level SEM, given the small variance attributable to the school level (see Table 3). Among the hypothesised mediators on school environments, pairwise correlation analysis was conducted to test how these variables are correlated with outcome (dependent) variables (see Table 4). Interestingly, only three measures related to structural quality—principals' qualifications (bachelor's degree or above), school has new mother tongue textbooks, and school has a library—were moderately correlated with the EGRA test scores, which ranged from 0.07 to 0.18. By contrast, the rest of school measures related to process quality (e.g., principals reviewed lesson plans or monitored teachers' instruction via classroom observation) had no significant or negative correlation with outcome variables.[99] The estimates from the *t-test* also indicated a statistically significant difference ($p < 0.01$) between preschool attendees and non-attendees in the three variables capturing structural quality. As explained, in Ethiopia, principals' qualifications and the availability of textbooks are positively associated with students' academic achievement (Piper, 2010; NEAEA, 2014; Woldehanna et al., 2008). Hence, I decided to include these three mediators in the analysis for each outcome variable.

To determine the best fitting model for the present data, I tested two different models. Model 1 first included preschool attendance as a direct predictor of three observed measures of school environment (i.e., principals' qualifications, the availability of textbooks, and school library), and children's EGRA test scores. Next, these three measures of school environments were included as direct predictors of children's EGRA test scores. To account for common sources of measurement error, error terms of the three variables representing school environment were allowed to be correlated. In Model 2—represented visually in Figure 3.4—I introduced the *latent* variable of school environments based on the selected observed variables. Model 2 included preschool attendance as a direct predictor of the latent measure of school environment and children's EGRA test scores, then this latent measure was included as a direct predictor of children's EGRA test scores. On the latent measure of school environment, I conducted a one-factor confirmatory factor analysis and confirmed that this shows an adequate model fit ($\chi^2(2) = 43.59$, $p < 0.001$; CFI = 0.95; RMSEA = 0.05). Taken together, Model 1 is the cumulative model in which the joint contribution of three observed school characteristics are included, while Model 2 builds up with the latent variable that can be regarded as a composite score of three indicators related to school environment.

[99] I tested some of the 'process quality' variables in the SEM model but did not achieve an adequate model fit.

Across all models, covariates were included to account for potential sources of selection bias. All covariates, including child age, gender, grade, paternal and maternal literacy, home reading resources, language of instruction, and urbanicity, have paths to preschool attendance, mediators, and students' outcomes. Analyses were conducted in Stata version 14.1. For all models, adequate model fit was indicated by a root mean squared error of approximation (RMSEA) of ≦ 0.06 (Hu & Bentler, 1999) and a comparative fit index (CFI) of ≧ 0.90 (Bentler, 1990). The traditional goodness-of-fit statistic (i.e., a nonsignificant chi-square) test was relaxed because the 'chi-square value can be overly influenced by sample size, correlations, variance unrelated to the model, and multivariate non-normality' (Kline, 2011, p. 201).

3.7 Results

3.7.1 Research Question 1: Variation in Preschool Influence by Sub-Groups[100]

Table 3.1 presents the results (as expressed by odds ratios) of the multivariate logit regression with interaction terms that relate to whether the relationship between preschool attendance and students' learning outcomes differed across the sub-groups of gender, urban-rural residency, father's literacy, mother's literacy, and having reading materials at home. The interaction terms were included separately for each model, which performed best in a log likelihood ratio-test, instead of specifying the model with multiple interaction terms.[101] Columns (5) and (6) present within-school differences by introducing the school fixed effects logit model for the EGRA 2016 cohort, which accounts for unobserved school characteristics that are fixed over time.[102]

Of all the interactions with preschool, gender was statistically significant in the 2016 cohort for the probability of being a non-reader, but not in the 2010 cohort (Models 1 to 4). This indicates that, in 2010, differences between preschool boys and preschool girls were similar to those between non-preschool boys and non-preschool girls, but the gains from preschool attendance were different for boys and girls in 2016. Namely, in 2016, preschool girls had a

[100] Descriptive statistics on sample characteristics were shown in Table 2.7.
[101] From the likelihood ratio (LR) test, the chi-squared value for the test and the p-value for a chi-squared value with two degrees of freedom were tested. In addition to the LR test, using Stata command *fitstat*, I compared log-likelihood, chi-square, R-squared (McFadden adjusted), AIC (Akaike information criterion), BIC (Bayesian information criterion), and variance of error terms across different models with single or multiple interaction terms. Consistent with the LR test, the model with single interaction terms were strongly supported, based on the difference in BIC.
[102] As mentioned in Chapter 2, the application of the school fixed effect to the EGRA 2010 cohort is not presented in this paper, due to little within-subject variance of preschool attendance in the 2010 cohort.

relative advantage over preschool boys (the odds of being a non-reader decreased by 29%), and this difference could be larger than the gap between non-preschool boys and non-preschool girls. Introducing the school fixed effect model in the 2016 cohort (Models 5 and 6) yielded consistent results, while the probability of being a proficient reader slightly increased among preschool girls.

Similarly, the null hypothesis that interactions between preschool and urban-rural residence (rural as reference) are different from zero cannot be rejected in 2010, but can be rejected in 2016. In 2010, the gains from preschool attendance were not significantly different between urban and rural children. In 2016, however, preschoolers living in rural areas had a significantly higher probability of becoming a non-reader (the odds increased by 55%) and lower probability of becoming a proficient reader (the odds decreased by 26%) than their peers living in urban areas. The within-school differences in the 2016 cohort (Models 5 and 6) were similar to the between-school differences (Models 3 and 4).

In the 2010 and 2016 cohorts, father's literacy strongly interacted with preschool in determining children's learning outcomes. In terms of the higher probability of being a proficient reader, the odds increased by 56 percent in 2010 and 37 percent in 2016 through the joint influence of preschool and father's literacy. In contrast, there is no significant interaction effects of the association between preschool attendance and mother's literacy. Regarding the measure of having reading materials at home, the interaction with preschool attendance was significant only for the 2010 cohort on the probability of being a proficient reader. If some interaction terms are statistically significant, the next questions will be, *What is the magnitude of interaction effects?* and *Are these effects large enough to matter?* To answer these questions, I will further explore the predicted probabilities and marginal effects of preschool attendance conditioned on the selected dimension: gender, urban-rural residence, and father's literacy.[103]

[103] The predicted probabilities and marginal effects of preschool attendance conditioned on mother's literacy or reading materials at home are available upon request.

Table 3.1. Interaction Effects of Preschool and Moderator Variables

	2010		2016		2016 (*school fixed-effects)*	
	(1) Non-Reader	(2) Proficient Reader	(3) Non-Reader	(4) Proficient Reader	(5) Non-Reader	(6) Proficient Reader
Odds ratio (SE)						
Pre	1.05 (0.21)	1.09 (0.15)	0.58*** (0.06)	1.35*** (0.14)	0.96*** (0.01)	1.00 (0.01)
Female	1.36** (0.17)	0.79** (0.07)	1.00 (0.09)	1.33*** (0.11)	0.85*** (0.1)	1.47*** (0.01)
Pre * Female	1.03 (0.17)	0.91 (0.17)	0.71** (0.11)	1.02 (0.13)	0.62*** (0.01)	1.14*** (0.01)
R-squared	0.13	0.08	0.21	0.11	0.10	0.05
Pre	1.32 (0.40)	0.95 (0.12)	0.35*** (0.07)	1.75*** (0.27)	0.52*** (0.01)	1.11***(0.02)
Rural	1.51 (0.42)	0.56** (0.13)	1.34** (0.16)	0.58*** (0.07)	-	-
Pre * Rural	0.74 (0.24)	1.14 (0.21)	1.55** (0.33)	0.74* (0.12)	1.57*** (0.04)	0.96* (0.02)
R-squared	0.13	0.09	0.21	0.11	0.11	0.06
Pre	1.50* (0.33)	0.79 (0.13)	0.51*** (0.07)	1.08 (0.14)	0.72*** (0.01)	0.84 (0.14)
Father’s literacy	0.64*** (0.07)	1.25** (0.14)	0.62*** (0.06)	1.25** (0.11)	0.66*** (0.01)	1.22** (0.10)
Pre * F’s literacy	0.56*** (0.12)	1.56** (0.30)	0.95 (0.16)	1.37** (0.21)	1.09***(0.02)	1.47** (0.26)
R-squared	0.13	0.08	0.20	0.10	0.21	0.05
Pre	0.89 (0.15)	1.04 (0.16)	0.47*** (0.05)	1.35*** (0.13)	0.76*** (0.01)	1.07*** (0.01)
Mother’s literacy	1.09 (0.11)	0.91 (0.11)	0.90 (0.09)	1.09 (0.09)	1.10*** (0.01)	1.06*** (0.01)
Pre * M’s literacy	1.46 (0.36)	1.01 (0.30)	1.11 (0.17)	1.05 (0.14)	1.01 (0.02)	1.00 (0.01)
R-squared	0.13	0.08	0.20	0.10	0.21	0.05
Pre	1.13 (0.24)	0.85 (0.12)	0.52*** (0.05)	1.41*** (0.13)	0.60*** (0.07)	1.07 (0.12)
R. M at home	0.58*** (0.08)	1.41*** (0.18)	0.50*** (0.05)	1.72*** (0.14)	0.51*** (0.06)	1.71*** (0.17)
Pre * R. M at home	0.77 (0.25)	2.04*** (0.57)	0.87 (0.14)	0.97 (0.13)	0.79 (0.17)	1.09 (0.17)
R-squared	0.13	0.09	0.20	0.10	0.21	0.05
All covariates	Yes	Yes	Yes	Yes	Yes (except region)	Yes (except region)
School-fixed	No	No	No	No	Yes	Yes
Observations (school)	9,121	9,121	8,332	8,332	7,461 (202)	8,020 (214)

Note: (1) Robust standard error in parenthesis; (2) Models 1, 2, 3, and 4 account for all covariates and include sampling weight; (2) Models 5 and 6 use school fixed effects and include sampling weights; number of schools in parentheses; (3) In Models 5 and 6, 23 groups (871 observations) in non-reader and 11 groups (312 observations) in proficient reader dropped because of all positive or all negative outcomes.

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

Source: EGRA Dataset 2010, 2016, USAID

Figures 3.1 to 3.3 present the average predicted probabilities and average marginal effects of the association between preschool and student outcomes differed by sub-groups. Note that, to compute these predictive values, I first fixed to each possible combination of the two variables of interest (e.g., preschool and gender), estimated the margins of each cell with observed values of other independent variables, and averaged these margins to obtain the single marginal effects. In the figures, predictive probabilities (left panel of each cohort) capture the difference between, for example, boys and girls, stratified by preschool attendees and non-attendees. Average marginal effects (right panel of each cohort) capture the discrete change of probabilities of being a non-reader or a proficient reader if he or she attended preschool.

Preschool and gender. Figure 3.1 displays the relations between preschool attendance and students' outcomes as a function of gender. Consistent with the findings shown in Table 3.1, the 2010 cohort shows that differences in predicted probabilities between preschool boys and preschool girls were similar to those between non-preschool boys and non-preschool girls. In turn, marginal effects were almost zero, regardless of gender. Conversely, in 2016, when preschool girls showed a relatively better performance than preschool boys (lower probability of being a non-reader), the gender gaps among preschoolers were larger than those between non-preschool girls and non-preschool boys. The figure with average marginal effects indicates that the gain in lower predicted value of non-reader was four percentage points larger for girls (13%) than for boys (9%) in 2016. The difference was small but statistically significant, at the 0.05 level (d = 0.04, SE = 0.02, $p < 0.05$).

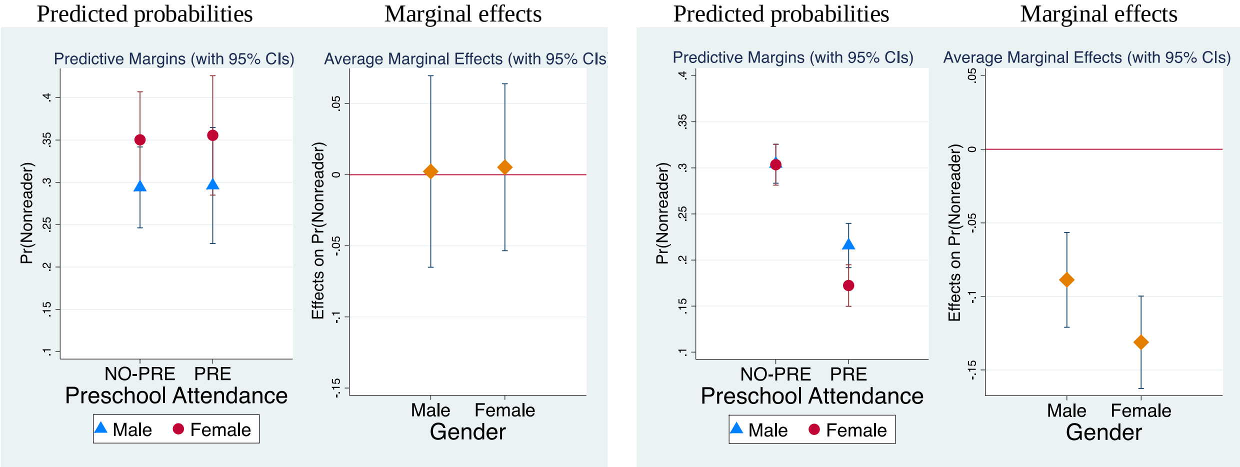
However, the preschool benefits of having a higher chance of being a proficient reader were similar for boys and girls across the 2010 and 2016 cohorts (Figure 3.1). Though overall marginal effects of the association between preschool and student outcomes increased from 2010 to 2016, there were no differential benefits by gender among each group of preschoolers and non-preschoolers. Moreover, we must pay attention to the interesting fact that the overall gender gap was fully reversed between 2010 and 2016. In 2010, before the preschool expansion, both preschool boys and non-preschool boys had a relative advantage (i.e., a lower chance of being a non-reader and a higher chance of being a proficient reader) over preschool girls and non-preschool girls; however, in 2016, girls outperformed boys and the gains were even greater for girls among those who attended preschool.

Figure 3.1. Predicted Probabilities and Marginal Effects of Preschool on Early Grade Reading Performance as a Function of Gender

(1) EGRA 2010: Preschool by Male-Female | (2) EGRA 2016: Preschool by Male-Female

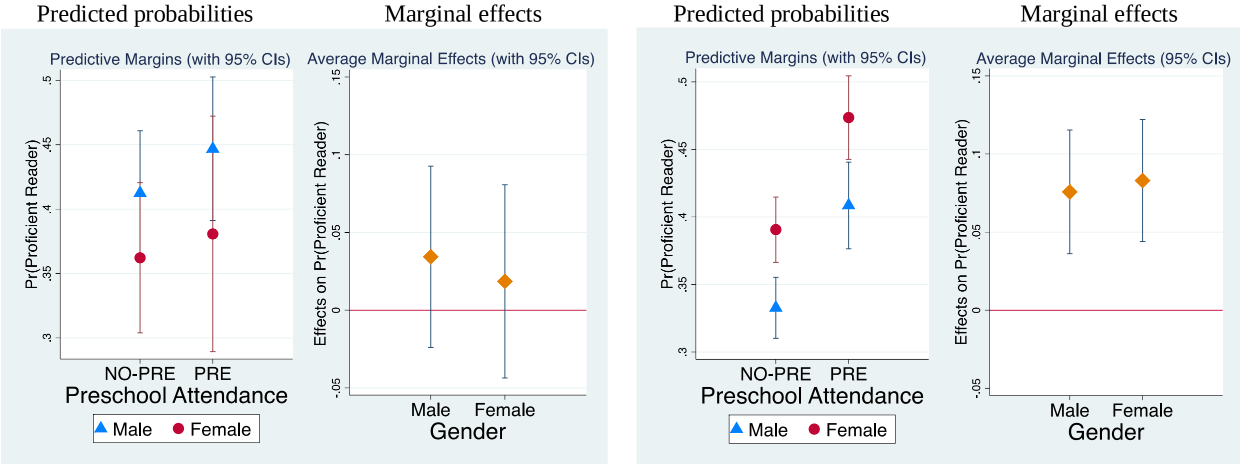
Non-Reader

Predicted probabilities | Marginal effects | Predicted probabilities | Marginal effects



Proficient Reader

Predicted probabilities | Marginal effects | Predicted probabilities | Marginal effects



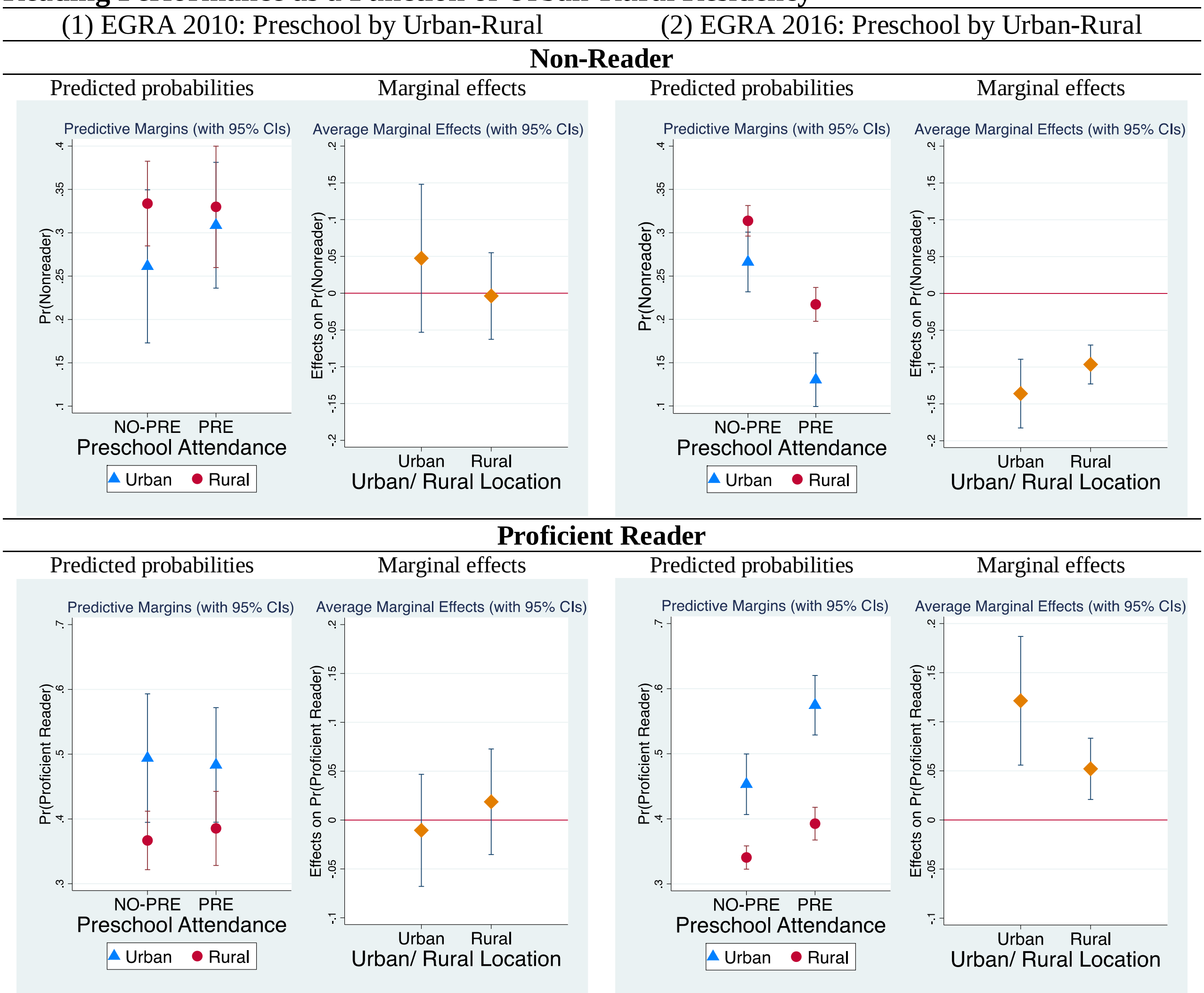
Note: Each cohort of EGRA 2010 and EGRA 2016 has two figures on predicted probabilities (Left) and marginal effects (Right).
Source: EGRA Dataset 2010, 2016, USAID

I further explored the gender gap based on preschool attendance, stratified by grade and region.[104] Although similar patterns were observed by Grade 2 and Grade 3, it is surprising that, in 2016, Grade 2 preschoolers showed a level of achievement similar to non-preschoolers at Grade 3, indicating that preschoolers were nearly one year ahead of non-preschoolers, regardless of gender. The patterns across the five regions between 2010 and 2016 varied but maintained similar gender gap patterns; for example, in SNNP and Amhara, after the large-scale expansion of preschool, girls who attended preschool made greater gains than any other group, including preschool boys, non-preschool boys, and non-preschool girls.

[104] The results are available upon request.

Preschool and urban-rural. I examined the extent to which the role of preschool differed by urban-rural residency. Figure 3.2 shows that, regardless of urban-rural residency, preschool benefits for having a lower chance of becoming a non-reader are similar in 2010 (d = 0.05, SE = 0.06, p > 0.1) and in 2016 (d = 0.04, SE = 0.03, p > 0.1). However, the gains from preschool, which favour children living in urban areas, are more pronounced in terms of the probability of being a proficient reader in 2016. Compared to there being no differential gains between urban and rural areas in 2010 (d = 0.03, SE = 0.04, p > 0.1), in 2016 the preschool benefits were greater for children living in urban areas than those in rural areas, and this difference was statistically significant at the 0.05 level (d = 0.07, SE = 0.04, p < 0.05).

Figure 3.2. Predicted Probabilities and Marginal Effects of Preschool on Early Grade Reading Performance as a Function of Urban-Rural Residency



Note: Each cohort of EGRA 2010 and EGRA 2016 has two figures on predicted probabilities (Left) and marginal effects (Right).
Source: EGRA Dataset 2010, 2016, USAID

Preschool and father's literacy. Finally, Figure 3.3 plots estimates from the model in which the indicator for preschool attendance interacted with father's literacy. In 2010, the benefits of preschool for having a lower probability of being a non-reader were greater for children with a literate father (d = 0.11, SE = 0.04, $p < 0.01$), whereas there was no differential benefit in 2016 (d = 0.01, SE = 0.03, $p > 0.1$). In terms of a higher probability of being a proficient reader, the benefits from preschool were greater for children with literate fathers. Notably, this pattern is statistically significant across two cohorts in 2010 (d = 0.10, SE = 0.04, $p < 0.05$) and 2016 (d = 0.08, SE = 0.03, $p < 0.01$). Taken together, the results show that some sub-groups—girls, children living in urban areas, children from households with literate fathers—benefited significantly more from preschool attendance. This provides suggestive evidence that the expansion of preschool could *reinforce* the learning gaps between advantaged and disadvantaged groups.

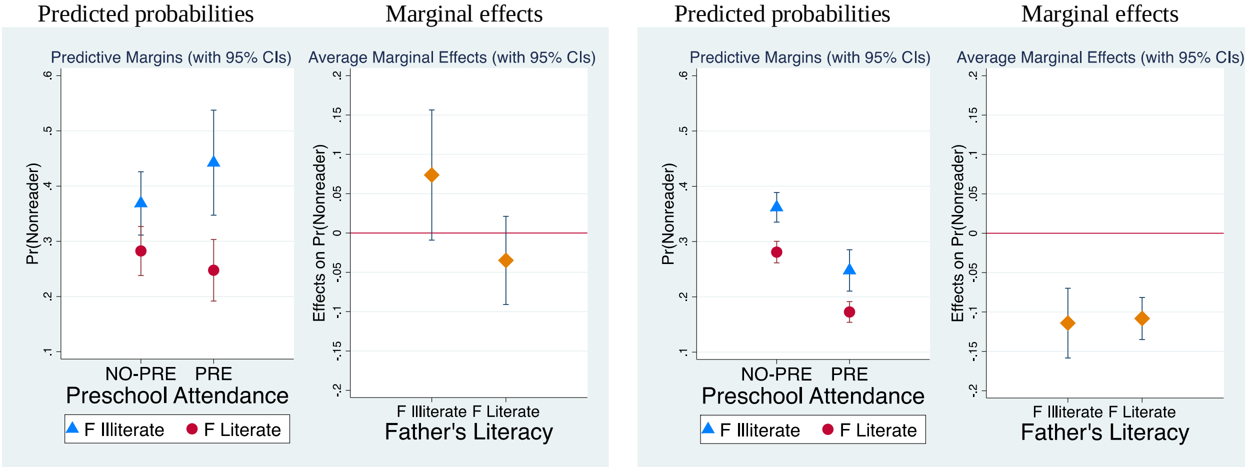
Lastly, to check the robustness of these results, I further explore the interaction effects of the association between preschool attendance and woreda (district) poverty index on student outcomes (Appendix B). The results showed that the benefits of preschool were particularly significant for children living in the richest woreda, whereas there were virtually no measurable benefits for children living in the poorest woreda. The results reaffirm that the large-scale expansion of preschool in Ethiopia may not contribute to reduce learning gaps induced by household poverty or location.

Figure 3.3. Predicted Probabilities and Marginal Effects of Preschool on Early Grade Reading Performance as a Function of Father's Literacy

(1) EGRA 2010: Preschool by Father's literacy | (2) EGRA 2016: Preschool by Father's literacy

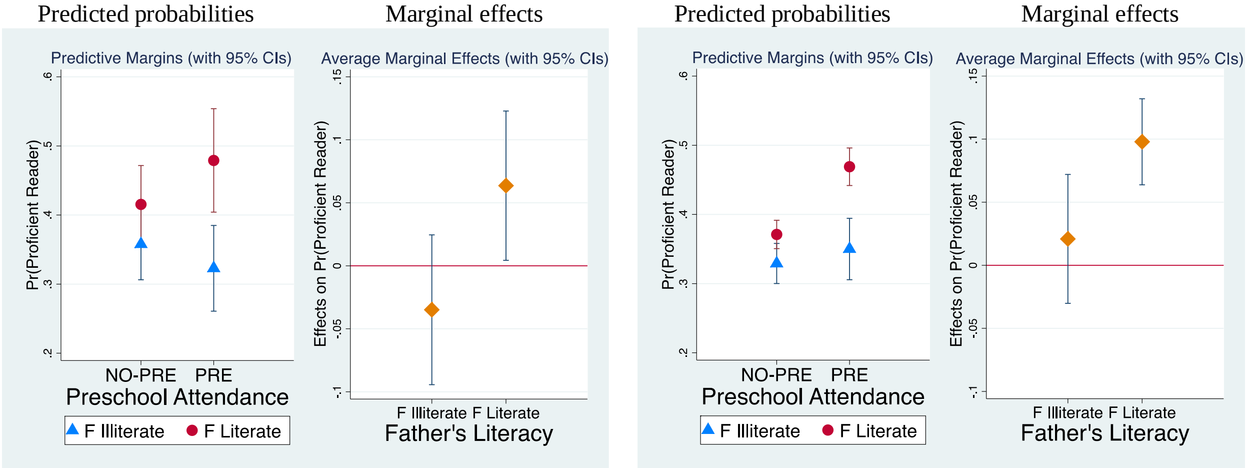
Non-Reader

Predicted probabilities | Marginal effects | Predicted probabilities | Marginal effects



Proficient Reader

Predicted probabilities | Marginal effects | Predicted probabilities | Marginal effects



Note: Each cohort of EGRA 2010 and EGRA 2016 has two figures on predicted probabilities (Left) and marginal effects (Right).
Source: EGRA Dataset 2010, 2016, USAID

3.7.2 Research Question 2: Mediating Role of Subsequent School Experience

1) Descriptive statistics

Descriptive statistics for school characteristics taken from EGRA 2010 and EGRA 2016 are presented in Table 3.2. There is overall improvement in the average information reported by school principals between 2010 and 2016. In 2016 there was a higher proportion of principals who hold a bachelor's degree or above (from 14% to 53%);[105] were responsible for classroom observation (from 23% to 51%); and who conducted oral examinations to monitor students'

[105] The rest of the school principals hold only a high school diploma.

reading performance (from 33% to 56%). In both cohorts, less than half the schools had a library for Grade 2 and Grade 3 students, while the distribution of mother tongue textbooks was nearly universal—more than 90 percent of schools received the textbooks.[106] About three-quarters of schools had teachers who participated in trainings for mother tongue reading instruction, which was provided by regional/local governments or development partners.[107] The ten listed school characteristics were not strongly correlated; among the three selected characteristics—principals' qualifications, textbooks, and having a library—correlations ranged from 0.06 to 0.26 (see Table 3.4).

Table 3.2. Descriptive Statistics: School Characteristics

School-level characteristics		EGRA 2010[108]		EGRA 2016		
		Mean	(SD)	Mean	(SD)	*Diff.*
(1)	Principal has bachelor's degree or above.	0.14	(0.35)	0.53	(0.50)	0.39
(2)	Principal received the training on early reading.	0.27	(0.44)	0.21	(0.41)	-0.06
(3)	Principal supports teachers for reading instruction.	0.83	(0.38)	0.88	(0.33)	0.05
(4)	Principal satisfied school's reading performance.	0.48	(0.50)	0.61	(0.49)	0.13
(5)	Principal reviewed lesson plans every week.	0.20	(0.40)	0.27	(0.44)	0.07
(6)	Principal responsible for classroom observation.	0.23	(0.42)	0.51	(0.50)	0.28
(7)	Principal conducted oral examination.	0.33	(0.47)	0.56	(0.50)	0.23
(8)	School has library for G2/G3 students.	0.41	(0.49)	0.42	(0.50)	0.01
(9)	School received (new) mother tongue textbook.	0.90	(0.30)	0.96	(0.21)	0.06
(10)	School has teachers who attended in-service training for MT instruction.	0.86	(0.35)	0.75	(0.43)	-0.11
Observation (Schools)		154		225		
Observation (Students)		5,843		8,332		

Note: (1) Of 225 schools in EGRA 2016, 200 responders were school principal (89%), 21 are deputy principal (9%), and 4 are other staff (2%). (2) About 65 percent of EGRA 2016 sample school has O-Class in 2014/15, according to matched data with EMIS.
Source: EGRA Dataset 2010, 2016, USAID

For child outcomes, I estimated the proportion of variance attributable to schools and children (nested in schools) by fitting a linear mixed model for each outcome with covariates and only a random intercept for school and a residual error for students (Table 3.3). Across all outcome measures of EGRA, nearly all (83% to 90%) of the variance in children's literacy was accounted for by differences across children, with approximately 10 percent of the variance

[106] The share of schools with a library is consistent with the national figure; 45 percent of government primary schools in Ethiopia have a library (MoE, 2015-2016); some schools have a library used by teachers only, coded as 'no library (0)'; new (revised) mother tongue curriculums and textbooks were introduced in 2014-2015.
[107] MT training implementers varied in 2010, organised by REB (77%), woreda (42%), cluster school or school (13%), or others (13%). In 2016, MT training was largely provided by USAID's READ-TA programme.
[108] In the 2010 EGRA, missing (or mis-coded) responses in the principal's questionnaire were high, at 36 percent. SEM did not apply to the 2010 EGRA cohort, due to the high rates of missing responses and low association between preschool and students' outcomes.

explained by differences in schools. Considering the relatively small intra-class correlations at the school level, I used a single-level SEM framework in the current analysis.[109]

Table 3.3. Intraclass Correlations (ICCs) of Dependent Variables

	Proportion of Variance	
Outcome Variables	Child	School
Oral Reading Fluency (ORF)	0.866	0.134
Letter Sound Recognition	0.832	0.168
Familiar Word Recognition	0.836	0.164
Invented Word Recognition	0.851	0.149
Reading Comprehension	0.888	0.112
Listening Comprehension	0.903	0.097

Source: EGRA Dataset 2010, 2016, USAID

[109] Additionally, the choice of the sing-level SEM model is related to the limited SEM functions in Stata. Stata provides 'sem' and 'gsem' (generalised sem) modelling options. Only gsem fits models to both single- or multilevel data, yet gsem does not provide key features such as goodness-of-fit statistics, model fit indicators, modification indices, and tests of indirect effects.

Table 3.4. EGRA 2016: Pairwise Correlations for Mediating and Dependent Variables

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)	(12)	(13)	(14)	(15)	(16)
(1) Qualification	1.00															
(2) Training	0.09***	1.00														
(3) Teacher support	-0.02	0.10***	1.00													
(4) Satisfaction	0.07***	0.08***	0.10***	1.00												
(5) Lesson Plan	-0.27***	0.13***	0.06***	-0.06***	1.00											
(6) Class observation	-0.16***	0.13***	0.10***	0.11***	0.37***	1.00										
(7) Oral test	-0.01	-0.21***	0.16***	0.04***	0.06***	-0.06***	1.00									
(8) Library	0.26***	0.00	0.12***	0.11***	-0.20***	-0.04***	-0.07***	1.00								
(9) MT Textbook	0.14***	-0.08***	0.13***	0.10***	-0.06***	0.04***	0.02	0.06***	1.00							
(10) Teacher Training	-0.02	-0.06***	-0.02	0.04***	0.00	-0.06***	0.01	0.04***	-0.07***	1.00						
(11) ORF	0.11***	-0.02**	0.03***	0.06***	-0.12***	0.05***	-0.02	0.15***	0.12***	-0.08***	1.00					
(12) Letter sounds	0.04***	-0.09***	0.02	-0.01	-0.13***	-0.02**	0.01	0.12***	0.09***	-0.03***	0.72***	1.00				
(13) Familiar words	0.16***	-0.01	0.05***	0.08***	-0.15***	0.03**	0.02	0.18***	0.14***	-0.06***	0.89***	0.74***	1.00			
(14) Invented words	0.15***	0.00	0.05***	0.07***	-0.13***	0.03***	0.01	0.13***	0.13***	-0.09***	0.86***	0.73***	0.89***	1.00		
(15) Reading compre.	0.07***	-0.03***	0.03***	0.03***	-0.10***	0.03**	-0.02**	0.12***	0.09***	-0.06***	0.83***	0.66***	0.79***	0.75***	1.00	
(16) Listening compre.	0.02	-0.09***	0.00	-0.03***	-0.13***	-0.09***	0.06***	0.04***	0.03**	0.06***	0.22***	0.36***	0.24***	0.21***	0.27***	1.00

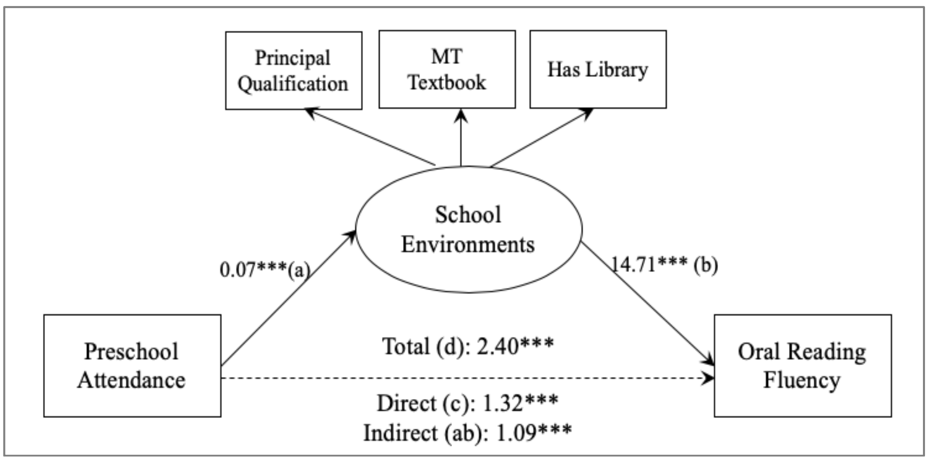
Note: *** p<0.01, ** p<0.05, *p<0.1
Source: EGRA Dataset 2010, 2016, USAID

2) Mediation analysis results

I conducted a mediation analysis to examine whether school-level characteristics mediated the relations between preschool attendance and students' early literacy outcomes. Overall, SEM Model 2, in which a latent variable was included, showed adequate model fit statistics: (with ORF; $\chi^2(22)$ = 323.69; CFI = 0.90; RMSEA = 0.04) (see Table 5). Compared to Model 2, Model 1 (which included three observed variables of school environment) showed a significantly poorer fit (with ORF; $\chi^2(3)$ = 722.48; CFI = 0.73; RMSEA = 0.17; SRMR = 0.02) (see Appendix Table D.10). Thus, I focus on presenting and interpreting the results of Model 2 as the final model.

From the SEM results of Model 2, standardised coefficients of the total, direct, and indirect effects on oral reading fluency (ORF) are presented in Figure 4. To ease the interpretation, I use the term 'effect' in the mediation analysis, where effect indicates 'association'. Preschool attendance positively influenced the mediator of schools' structural quality (β = 0.07, SE = 0.10, p < 0.001). Preschool attendance was also a positive predictor (direct effect) of ORF scores in the model (β = 1.32, SE = 0.48, p < 0.001), after controlling for child age, gender, grade, paternal literacy, maternal literacy, home reading resources, language of instruction, and urbanicity. In addition to the direct effects, I assessed the indirect effects among the study variables. The indicator of schools' environments (structural quality) was found to be a statistically significant mediator (β = 1.09, SE = 0.18, p < 0.001).

Figure 3.4. SEM Model on Preschool, School Environment, and Oral Reading Fluency



Note: (1) Total Effects (d, 2.40) = Indirect Effects (a*b, 0.0737687 * 14.71128) + Direct Effects (c, 1.32)
*** p<0.01, ** p<0.05, *p<0.1
Source: EGRA Dataset 2010, 2016, USAID

The results of the SEM model for all six outcome variables are presented in Table 3.5, including students' test scores as measured by ORF, letter sounds, familiar words, invented words recognition, reading and listening comprehension. In three out of six outcomes—letter sounds recognition, reading, and listening comprehension—the model fit was marginally unsatisfactory (RMSEA = 0.87 to 0.88, compared to the recommended level of ≧ 0.90). The model with adequate fit statistics—familiar words and invented words recognition—showed that preschool attendance was a positive predictor (direct effect) of familiar words recognition (β = 1.25, SE = 0.48, p < 0.001) and invented words recognition (β = 1.41, SE = 0.34, p < 0.001). In this model, the indicator schools' environments (structural quality) was found to be a statistically significant mediator (β = 1.57, SE = 0.25, p < 0.001 for familiar words recognition; β = 0.89, SE = 0.18, p > 0.001 for invented words recognition) for the link of preschool attendance with students' outcomes.

I further assessed the indirect relations among the study variables. For mediation effects analysis, no single index appears to be a viable mediation effect size measure (Wen & Fan, 2015). Given this limitation, I followed the recommendations of Sobel (1982) to use (1) the proportion of total effect that is mediated (i.e., total effect explained by the indirect effect); and (2) the ratio of the indirect effect to direct effect (R_m statistic) as a proxy measure for the magnitude of mediation effect. Specifically, R_m < 1, R_m=1, or R_m> 1 indicates that the indirect effect is smaller than, equal to, or larger than the direct effect, respectively. As shown in Table 3.5, the indirect path from preschool attendance to students' ORF score via schools' structural quality explained 45.0 percent of the total association between preschool attendance and ORF score (β = 2.40), and the ratio of the indirect effect to direct effect is 0.82. In addition, the indirect path mediated by schools' structural quality explained 55.7 percent of the total association between preschool attendance and familiar words test score (β = 2.82), with the ratio of the indirect effect to direct effect at 1.26; and 38.7 percent of the total association between preschool attendance and invented words test score (β = 2.40), with the ratio of the indirect effect to direct effect at 0.63. Notably, the contribution of the indirect effects is either close to or outweighs the direct effects in oral reading fluency (R_m = 0.82) and familiar word recognition (R_m = 1.26). In this model, the mediator of schools' structural quality explained about half of the total effects for the association between preschool attendance and students' outcomes.

Taken together, the results of the SEM suggest that the estimated path of preschool attendance on students' early reading performance was partially accounted for by good school environments. This analysis is still exploratory, given that a limited measure of school characteristics was used and the multi-level structure of school was not fully counted by the SEM model. Nevertheless, the findings call for more attention to the subsequent schooling experience that could affect the sustained benefits of preschool. Ensuring a smooth transition between pre-primary and primary education is the critical foundation for students' lifelong educational trajectories.

Table 3.5. EGRA 2016: SEM Standardised Coefficients and Direct and Indirect Effect

	(1) ORF	(2) LS	(3) FW	(4) IW	(5) RC	(6) LC
Path Coefficients **(S.E.)**						
Preschool Attendance → School Structural Quality (Environments) → EGRA Outcome						
Total effects	2.40*** (0.46)	8.54*** (0.70)	2.82*** (0.46)	2.30*** (0.33)	3.32*** (0.60)	4.22*** (0.62)
Direct effects	1.32*** (0.48)	8.01*** (0.72)	1.25*** (0.48)	1.41*** (0.34)	2.39*** (0.62)	4.20*** (0.63)
Indirect effects	1.08*** (0.18)	*0.53**** *(0.17)*	1.57*** (0.25)	0.89*** (0.14)	0.93*** (0.18)	0.02 (0.09)
% of total effect mediated	45.0	6.2	55.7	38.7	28.0	0.5
% of total effect unmediated	55.0	93.8	44.3	61.3	72.0	99.5
Ratio of indirect effect to direct effect (R*m*)	0.82	0.07	1.26	0.63	0.39	0.00
Model Fit						
Chi-Square (*DF*)	323.7(22)	340.4(22)	318.7(22)	287.6(22)	311.7(22)	249.5(22)
CFI	0.90	0.88	0.90	0.90	0.88	0.87
RMSEA	0.04	0.04	0.04	0.04	0.04	0.04
SRMR	0.02	0.02	0.02	0.02	0.02	0.02
R-Square	0.18	0.20	0.19	0.16	0.18	0.12

Note: (1) The resulting structural coefficients (standardised regression coefficients) describe the direct and indirect effects, based on students' test scores measured by letters/words per minute in Model 1 to 4, and percentage of correct answers in Model 5 and 6; (2) Standard errors are in parentheses; (3) ORF oral reading fluency; LS Letter sound; FW familiar word recognition; IW invented word recognition; RC reading comprehension; LC listening comprehension; (4) DF: Degree of Freedom; (5) CFI: Comparative Fit Index; (6) RMSEA: Root Mean Square Error of Approximation; (7) SRMR: Standardised Root Mean Square Residual.
*** p<0.01, ** p<0.05, *p<0.1
Source: EGRA Dataset 2010, 2016, USAID

3.8 Discussion and Future Direction for Research

The present study is the first to employ two large and representative datasets to examine the relationship between preschool attendance and students' learning outcomes across child, family, and school characteristics, and whether these patterns are different before and after the early learning reform in Ethiopia. The primary aims of this study were specifically to understand (1) whether the benefits of preschool were moderated by sub-group as defined by gender, urban and rural location, parental literacy, and home reading resources; and (2) whether the relation between preschool and students' outcomes was mediated by subsequent schooling environments, as captured by primary school characteristics (i.e., principals' qualifications, availability of current textbooks, and a school library). The present study focused on empirically testing hypotheses derived from bioecological theory, which imply that the relationship between preschool and learning outcomes depends on characteristics of children's multi-layered environments (family, school, community, and policy).

Differential benefits by gender. As for gender, it is encouraging that the findings showed that the benefits of preschool attendance were particularly pronounced for girls after the expansion of O-Class in Ethiopia. In the region-wide learning assessments, such as the SACMEQ or PASEC in Africa, girls performed as well as or better than boys on reading assessment (Bashir et al., 2018).[110] However, in Ethiopia there were significant gender differences in learning that favoured boys, which seemed to be entrenched across the five rounds of the Ethiopia National Learning Assessments (NEAEA, 2016; World Bank, 2016). Interestingly, however, the present study showed that the average achievement gap in early grade reading between girls and boys was reversed between 2010 and 2016: boys outperformed girls in 2010, whereas girls not only outperformed boys in 2016 but even obtained greater benefits from preschool than boys. While only a few studies in LMICs have addressed the gender gap, most of them found no significant gender differences in the benefits of attending preschool on academic performance (Berlinski et al., 2009; Bastos et al., 2017; Biethenbek et al., 2017; Agirdag et al., 2015).

Findings from the present study add new evidence that could open policy dialogue and inform future research into patterns of the gender gap in the educational trajectories. There also is a need for further exploration of why the achievement gap between boys and girls was reversed across time, as well as between the early grades and the upper grades. Although girls started to perform well in reading in Grades 2 and 3, the National Learning Assessments for students at Grades 4, 8, 10, and 12 consistently showed that boys performed significantly better than girls in various subjects.[111] Moreover, it should be noted that the gender gap tends to be compounded

[110] SACMEQ stands for Southern and Eastern Africa Consortium for Monitoring Education Quality; PASEC stands for Programme d'analyse des systèmes éducatif de la CONFEMEN, meaning Programme for education systems analysis conducted in Francophone (Western) Africa.
[111] To illustrate, in Grade 4, the achievement of boys on average was higher than that of girls by 3.02 percent. The mean difference in four subjects between boys and girls ranged from 2.03 percent (reading) to 4.84 percent

by other sources of inequality, such as urban-rural location. In Ethiopia, the gender gap is more pronounced in rural than in urban contexts (Piper, 2010; World Bank, 2016), thus special attention should be given to girls in rural areas, who perform the least well.[112]

A future study should specifically examine the preschool effect by gender and the possible causes of this differential effect—that is, whether the curriculum challenges girls and boys in the same way, and whether girls and boys have the same academic needs across the educational trajectories and based on where they live. A future study could (a) analyse the gender gap using a more recent, nationally representative sample to see if there is a concurrent finding; (b) study whether boys and girls enter preschool with similar basic literacy and numeracy skills and academic needs; (c) study teachers' and parents' expectation regarding preschool-age boys and girls, as well as any cultural values around gender roles; and (d) explore whether the national curricular guides are gender sensitive. A study reflecting these aspects should inform efforts to improve ECE standards and national curricular guides to better serve both boys and girls.

Differential benefits by urban-rural residency, parental literacy, and home reading resources. The results of this study suggest that the benefits of increased access to preschool were not particularly pronounced for students from disadvantaged backgrounds, those living in rural areas, or those with illiterate fathers; however, the gains were larger for their peers living in urban areas or with literate fathers. The preschool benefits also appeared to be smaller or not significant for children living with illiterate mothers or having fewer books at home, and those who may have a less stimulating home learning environment. These findings do not 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 the U.S. and other high-income countries (e.g., Burger, 2010; Magnuson & Duncan, 2017). In fact, there is little support for the compensatory hypothesis among prior studies conducted in Sub-Saharan Africa, especially the cases using a large, nationally representative sample. In South Africa, after the mass expansion of pre-primary classes (also known as R-Class) over a decade, preschool attendance further extended the advantages of children from more affluent

(mathematics) and is significant (NEAEA, 2016). Similarly, the achievement gap between boys and girls is 4.33 percent in Grade 10 and 8.6 percent in Grade 12 (NEAEA, 2014).

112 However, this is not the case of the current analysis, which used EGRA 2016: girls outperformed boys in 2016, and rural boys performed the least well.

background (Berg et al., 2015). There were contrasting gains from preschool that favoured schools with rich resources and those located in high-performing regions—that is, there were no measurable preschool benefits for the poorer schools *but* considerable benefits for the wealthiest schools (Berg et al., 2015).[113] In Kenya and Tanzania, while there was a positive influence for all preschool attendees, there was no statistically significant difference across sub-groups by household wealth, urban and rural residency, and mother's education level (Bietenbeck et al., 2017).

The findings of the present study provide suggestive evidence that the large-scale expansion of preschool can *reinforce* the learning gaps between children from wealthier and poorer backgrounds. This pattern is rather relevant to the complementary hypothesis, which is also known as 'Matthew effects', wherein an initial advantage leads to cumulative advantage over time, thereby creating a virtuous cycle of continuous gain (Walberg & Tsai, 1983; Stanovich, 1986). The Matthew effects often are described as 'the rich get richer and the poor get poorer'. This could provide a partial explanation for the widening achievement gap, as children who grow up in a high-quality home learning environment profit more from the enriched early learning environment in preschool than their peers who grow up in a low-quality home learning environment. This has further social implications, in that a large-scale expansion of preschool without careful planning, targeted resources, and ongoing monitoring may have the negative consequence of exacerbating the educational inequality (Engle et al., 2011) that emerged before children entered primary school and increased through subsequent formal schooling.

Another possible explanation for this alarming pattern could be the poor quality of preschool, particularly in marginalised communities. In fact, O-Classes in Ethiopia, which are regarded as a rural phenomenon, have suffered from multiple challenges in delivering quality early learning programmes, including untrained or poorly qualified teachers, a lack of teaching and learning materials, large class size, and a lack of quality monitoring and supervision, among others (Teferra & Hagos, 2016). Large regional variations in financial and human resources for pre-primary education also risk delivering minimal benefits to those most in need. The current analysis was not able to test this hypothesis, due to the absence of quality measures of preschool

[113] In South Africa, using the data from the Annual Education Assessment, schools are categorised into five groups (quintiles) based on the relative wealth of their surrounding communities. Preschool effects have compared the two-lower quintiles (poorest) to the highest quintile (richest).

in the available dataset; this is a prominent area for future research. Future research must strive to disentangle whether the differential benefits of preschool are driven by differences in preschool quality or by any other factors attributed to family, community, and school characteristics. Despite the potential solutions for persistent low learning levels offered through increased access to preschool, the best ways to achieve equitable access and meaningful learning gains for all children should be further investigated.

The present study offers some insights for policymakers, including that expanding coverage of pre-primary education in isolation is not enough to reduce learning inequality for young children in Ethiopia. More effective ECE policies are needed to ensure that children, regardless of their geographical constraints or household poverty status, are able to learn and benefit from the educational experiences provided. Despite the government's ambition to provide universal pre-primary education by 2020 (at least one year prior to entering primary school), the expansion patterns and implementation strategies have not yet determined which compensatory measures can extend the advantage of preschool to most marginalised children. If O-Classes in rural areas are not meeting the goal of providing the conditions necessary for poor children to learn and to enter primary school as well prepared as their more affluent peers, the government should recalibrate this policy. Future studies can explore which needs of disadvantaged families in rural areas are not being met by current ECE provision in order to help redesign the ECE policy. Effective policy should ensure equity for children from disadvantaged groups as one way to compensate for the different circumstances in which children were born.

Subsequent school characteristics as mediators. Finally, to determine which factors mediated the relation between preschool and learning outcomes, the present study tested a hypothesis that subsequent school environments play a role in sustaining preschool benefits. The school characteristics I considered in the current analysis, including principals' qualifications, the availability of textbooks, and the school having a library, partially mediated the relation between preschool attendance and students' academic achievement when I treated these as the latent variables of a school's structural quality. My findings align with previous work suggesting that attending high-performing or better-resourced schools mediated the link between preschool and academic performance at Grade 5 (Curenton et al., 2015) and educational attainment at age 20 (Reynolds et al., 2004). The results of the previous studies consistently indicated that the indirect effects of subsequent school experience accounted for one-fifth to one-third of the total effects for the association between preschool and student outcomes, which are small but statistically significant. Exploring primary school characteristics is particularly germane to the Ethiopian context, where nearly three-quarters of preschools are attached to a government primary school through O-Class. Due to the O-Classes being nested within the primary schools, the learning conditions between the two shared many features, such as infrastructure (e.g., classrooms, playground, toilets), school management and leadership under the primary school principal, and teachers recruited from the same pool of applicants.[114]

The mediation analysis in the present study is exploratory in nature, given the limited measures of school characteristics that are most central to learning outcomes. Nevertheless, the results of this study highlight the need for policy measures that push to enhance the primary school environment as a continuum of the quality preschool experience. Consistent with the dynamic skill formation models illustrated by ‘skills beget skills’ (Cunha & Heckman, 2010), earlier gains from preschool are likely to be maintained if children are subsequently exposed to conducive learning environments. There is a need for future research that precisely measures the quality of early grade schooling and identifies the inputs and features of school that are critical to the preservation of preschool effects (Bailey et al., 2017; Bassok et al., 2018; Phillips et al., 2017).

In recent years, the primary school environment in Ethiopia has changed substantially through the multi-phased reforms that emphasise meaningful gains in students’ learning outcomes (MoE, 2015; World Bank, 2017). Research is needed to assess the effects of these changes on students’ academic performance and to examine whether they have jointly influenced the sustained benefits of preschool. Future research also should highlight the importance of a smooth transition from pre-primary to primary education and how to ensure the best alignment of curricula, professional development, and teacher-child interaction across the foundational grades, an area which is currently under-researched in the LMIC context.

[114] Generally, teacher quality tends to be lower in O-Class than Grade 1. Qualitative interviews (personal) suggest that primary schools are likely to assign low-performing teachers to O-Classes (like a punishment). Moreover, due to irregular compensation and poor working conditions, it is not possible to attract competent candidates. Often, community workers without any experience or training in early childhood education lead O-Classes as a facilitator.

3.9 Limitations

Given the current analysis using the data constructed in the previous chapter, similar limitations apply to this chapter, including the limited measures of students' SES or household income levels, the lack of details on children's preschool experiences (e.g., preschool type, duration, and quality), and the narrow focus on early literacy skills (see Section 2.10). The first limitation here is the blunt measure of primary school characteristics provided by the EGRA. The school-level indicators used in the current analysis relied on principals' self-reports. No information was collected via class observation by a third party, that could have allowed researchers to cross-check the accuracy and objectivity of the information. Moreover, the indicators could have been improved by reflecting the recent changes in primary school environments. Many school-level variables collected by EGRA 2010 and EGRA 2016, such as principals' support for early grade reading, were unlikely to show a strong correlation with students' academic performance. Rich measures for primary school characteristics would provide insights about the role of sustaining environments that could leverage early learning gains. The second limitation is the single-level SEM mediation model used in the present study. Since I focused on the total indirect effects through the school-level mediator, this model could not distinguish the between- and within-school components of the indirect effects. Future study can consider using a multi-level SEM mediation model (Preacher et al., 2010) to examine the between- and within-school components of the indirect effects for a Level 1 treatment variable (preschool attendance) through a Level 2 mediator (primary school characteristics) on the Level 1 outcome (i.e., 1-2-1 multi-level mediation model).

Lastly, the current study used the conventional terms with 'effect' for specific empirical approaches, such as marginal effects in logit regression model and total, direct, and indirect effects in the SEM model. These estimates demonstrated the association between preschool attendance and students' learning outcomes that adjust for potential sources of selection bias but do not imply causal inference, as it is not possible to completely rule out differential selection into preschool that results from unobserved confounds.

3.10 Conclusion

The present study provides important knowledge for policymakers and practitioners in Ethiopia who are interested in ECE as an instrument to 'improve equity at the point of entry to the education system' (MoE, 2015). The results of this study highlight that the benefits of preschool related to gender, urbanicity, and parental literacy are not equally distributed after the large-scale expansion of O-Class in ways that have implications for ECE provision, specifically the ability to provide equitably effective educational opportunities. As a result, the rapid expansion of O-Class may not be achieving the equity goal of improving outcomes for children from diverse backgrounds. Immediate action must be taken to ensure that pre-primary education policies, funding, and curricula reflect this important diversity and allow flexibility in the provision of ECE to meet the needs of children from disadvantaged families and communities. The results of the current study suggest in particular that more intensive, high-quality early interventions are needed in rural communities and for those who have a less stimulating home learning environment. Future research is needed to identify additional contextual characteristics that support and/or undermine the effectiveness of ECE and to closely monitor whether preschool delivers beneficial effects for the children most in need.

4 CHAPTER 4 – Sustained Preschool Influence on Students' Learning Outcomes in Adolescence: Longitudinal Evidence from Ethiopia

4.1 Introduction

While the previous two chapters have looked at the relation between preschool and student academic outcomes in the early primary grades, this chapter focuses on the longer-term benefits of preschool on student educational outcomes in adolescence. Ensuring that young children benefit from their early learning experiences is essential to building a productive and equitable society. As a critical period of growth in cognitive, social, and emotional skills, investment in early childhood lays the foundation for lifelong learning and determines a child's ability to shape a positive academic trajectory, which in turn improves their later life outcomes (Shonkoff & Phillips, 2000; Knudsen et al., 2006). Pioneering experiments of early childhood programmes (Abecedarian: Campbell et al., 2012; Chicago Parent-Child Centres: Reynolds et al., 2011; Perry Preschool: Schweinhart et al., 2005) and theories from the economic and developmental literature on skill formation (Bailey et al., 2017; Cunha & Heckman, 2007) have demonstrated the compelling long-term impact early childhood interventions have on educational attainment, health, and labour market outcomes.

Early childhood interventions, policy in particular, are being promoted as cost-effective measures to rectify early life disadvantages, since this is the age when persistent development gaps and deficits occur (Heckman, 2006; Heckman, Pinto, & Savelyev, 2013). There are dynamic complementarities associated with investment in the early years, especially for disadvantaged children, which can make investment in subsequent years more productive (Cunha & Heckman, 2007). A number of cost-benefit analyses have revealed that early childhood interventions generate high-value benefits for society, higher than most investments later in the life cycle. The projected expansion of preschool in LMICs from serving one-quarter to one-half of young children is estimated to return US$6.4 to US$17.6 for each dollar invested (Engle et al., 2011); this is similar to the Perry Preschool Programme in the U.S., which has a benefit-to-cost ratio ranging from 7:1 to 13:1 (Belfield et al., 2006; Heckman et al., 2010). Reflecting this promise, early childhood education systems in LMICs have been burgeoning in recent years (Richter et al., 2017; Vargas-Barón, 2015), as governments, multinational organisations, and NGOs pursue these aspirational goals across highly diverse systems.

Despite the extensive literature on the medium- to long-term contribution of ECE in high-income countries (for a review, see Ruhm & Waldfogel, 2012; McCoy, Yoshikawa, et al., 2017), little is known about whether preschool benefits to child outcomes in LMICs persist into later years. In fact, no long-term ECE studies on adulthood outcomes such as employment and earnings exist in the low-income context, and only a handful of studies—none of which uses longitudinal data or experimental designs—showed the benefits preschool had on students' outcomes in secondary education and above:[115] school enrolment at age 7-18 in rural North India (Hazarika & Viren, 2013); school progression and cognitive skills at age 13-16 in Kenya and Tanzania (Bietenbeck et al., 2017); secondary education completion at age 17-18 in Ethiopia (Woldehanna & Araya, 2017), and educational attainment at age 18-29 in Egypt (Krafft, 2015).

A major puzzle in the extant literature is the *fadeout* of the initial academic gains from preschool, which have been observed in many experimental studies in the U.S (Bailey et al., 2017). This phenomenon, which can be more accurately described as a 'convergence' that reflects preschool attendees' *fadeout* and non-attendees' *catch-up,* typically occurred over the course of primary school (Yoshikawa et al., 2013). While a robust body of research has attempted to reconcile this early convergence by demonstrating that the benefits of preschool persisted into adulthood (Deming, 2009; Ludwid & Miller, 2007), the ability of ECE to improve children's educational outcomes in middle childhood and adolescence remained uncertain in recent ECE initiatives. Understanding the short-term benefits of preschool on a child's educational trajectory regarding—defined as the benefits during the preschool years or into the early grades—and the medium-term benefits of preschool—defined primarily as the benefits during the secondary school years—has important theoretical and policy implications, as it could help identify critical periods in the life course that are central to formulating policies that boost or at least maintain the beneficial effects of preschool. Importantly, the degree to which early education represents a wise investment is determined not only by improved school

[115] There is a well-known randomised experiment of early childhood interventions in Jamaica, which revealed significant long-term labour market returns (e.g., 25% increase in earnings at age 22) (Gertler et al., 2014). However, this experiment focused on a home-based early childhood stimulation intervention (i.e., weekly home visits for teaching parenting skills and better mother-child interaction), which is different from the focus of the present study's focus on centre-based ECE programmes.

readiness but also by how well these early gains are sustained over time (Magnuson et al., 2007).

To fill this gap, the present study examines how the relation between preschool attendance and student outcomes evolves from early childhood to adolescence. Using a longitudinal data from the Young Lives Study in Ethiopia, I assessed whether preschool attendance is persistently predictive of students' educational outcomes by age 15, the age at which most students are transitioning to secondary school. Specifically, applying matching procedures to a uniquely rich dataset which tracked children over 15 years, I compared socio-demographically similar preschool attendees and non-attendees on a range of educational outcomes, including academic achievement in receptive vocabulary, mathematics, and language, and the highest grade achieved at ages 8, 12, and 15. In addition to exploring these overall associations, I examined the extent to which the links between preschool attendance and later outcomes differed based on three theoretically guided dimensions: (1) the socio-demographic characteristics of child and family; (2) the characteristics of the preschool experience; and (3) subsequent schooling environments in primary education.

This study's focus on the cumulative influence of preschool attendance in a low-income context and its pathway through adolescence is a particularly valuable addition to the literature. Extended from previous studies on ECE in Ethiopia (Woldehanna, 2016; Woldehanna & Araya, 2017; Vandemoortele, 2018), I used a more comprehensive measure of student outcomes up to age 15 and considered the differential influence of preschool in relation to child, family, and preschool characteristics. Furthermore, this study takes into account both the direct and indirect effects of the association between preschool and student outcomes relative to subsequent schooling experience, thereby providing a more in-depth understanding of how and whether preschool benefits persist in terms of students' outcomes in adolescence. Given that early and later learning are complementary (Cunha & Heckman, 2007), the present study will enhance understanding of the contribution early learning makes in setting a path toward success both in primary school and later in life.

The rest of this chapter is structured as follows: I summarise the relevant literature in Section 2, set out the purpose and research questions of the present study in Section 3. I describe the data in Section 4 and provide the empirical methods used in Section 5, followed by the descriptive statistics and model fit in Section 6. I provide the results from the analysis in Section 7, discuss the findings and limitations in Sections 8 and 9, and conclude in Section 10.

4.2 Relevant Literature

4.2.1 Empirical Evidence on the Medium- and Long-Term Effects of ECE

ECE in general is targeted at fostering children's foundational skills that enable them to progress along a successful educational path (Heckman et al., 2013). These skills include cognitive skills in early literacy, numeracy, and language, and non-cognitive skills in self-regulation, motivation, and social competence. Educational outcomes in middle childhood or adolescence are regarded as more distal targets of ECE (McCoy, Yoshikawa, et al., 2017). For the present literature review, I focus on the medium- to long-term effects ECE has on students' educational outcomes. Following Rhum and Waldfogel (2012) and McCoy, Yoshikawa, et al. (2017), I define medium-term effects as those measured between ages 10-19, such as test scores, grade retention, and completion of secondary education, and long-term effects as those measured above age 20, such as final grade completed, employment, and earnings.[116] To avoid any confusion, it should be acknowledged that some meta-analyses used a different definition of long-term effects: Nores and Barnett (2010), for example, defined short-term as up to age 7 (roughly the beginning of compulsory formal schooling) and long-term as age 7 and above. Similarly, Rao et al. (2014) defined long-term as at least six months after the intervention was completed. In the current review, I first describe studies that assessed medium- to long-term effects in high-income countries, then discuss studies that assessed medium-term effects in LMICs. Variables that moderate or mediate the effects of ECE are explained in each study, if available.

High-income countries. In the U.S., three well-known model programmes—Abecedarian, Chicago Parent-Child Centres, and Perry Preschool—provide strong evidence on the medium- to long-term effects of ECE. Although cognitive test scores for children attending preschool and those not attending preschool typically converge over the course of primary school (Barnett, 2008), all these programmes were likely to yield sizable benefits on educational, behavioural, and health outcomes that persisted into adulthood (Campbell et al., 2012;

[116] World Health Organisation (WHO) defines an adolescent as any person between ages 10 and 19.

Reynolds et al., 2011; Schweinhart et al., 2005).[117] To illustrate, Temple and Reynolds (2007) documented that participation in the three ECE programmes led to a significant reduction in special education placement and grade retention, and to increases in high school completion and college attendance rates. Moreover, Perry Preschool and Chicago Parent-Child Centres participants were significantly less likely to turn to crime, and the Perry Preschool programme led to significantly higher rates of employment and earnings. Most recently, a meta-analysis of 22 experimental and quasi-experimental studies in the U.S. conducted between 1960 and 2016, including the three above, revealed that, on average, participation in ECE led to statistically significant reductions in special education placement (d = 0.33 *SD*, 8.1 percentage points) and grade retention (d = 0.26 *SD*, 8.3 percentage points), and to an increase in high school completion rates (d = 0.24 *SD*, 11.4 percentage points) (McCoy, Yoshikawa, et al., 2017). These results emphasised the effectiveness of ECE for reducing education-related expenditures and promoting child well-being.

In the U.K., Goodman and Sianesi (2005) investigated the benefits of ECE on a wide range of short- to long-term outcomes, including cognitive achievement from age 7 to 15, socialization, and adult outcomes on employment and earnings at age 33. The authors found positive influences of preschool attendance before age 5 on students' test scores, which attenuated in magnitude but remained statistically significant up to age 15, while the effects on socialization were more elusive. In adulthood, preschool education led to significant increases in the probability of obtaining qualifications and of being employed at age 33. Using more recent data from the Longitudinal Study of Young People in England, Apps, Mendolia, and Walker (2013) found that preschool attendance moderately improved cognitive tests scores at ages 11, 14, and 16, while there were no significant gains in psychological well-being, crime involvement, and health behaviours. The gains in cognitive skills were especially noticeable for girls and for children from disadvantaged socioeconomic backgrounds, which implies that ECE helps to reduce learning inequalities up to adolescence.

[117] As an example of convergence, the Perry Preschool project found that the cognitive advantages of preschool tended to decline over time, with the control group generally catching up cognitively when they entered kindergarten. The Abecedarian Project found similar declines in IQ over time, with effect sizes decreasing from 0.75 standard deviations (*SD*) at age 4 to 0.33 *SD* at ages 15 and 21 (Barnett, 2008).

Recently, further rigorous evidence has emerged from outside the U.S. and the U.K.[118] Using spatial and time variations in preschool availability in Norway, Havnes and Mogstad, (2011) found strong beneficial effects of preschool on subsequent educational attainment, including more years of schooling, a higher rate of college attendance, and a lower rate of high school dropout, and on adult outcomes including labour participation and reduced dependence on welfare. Preschool benefits were greatest for children with less-educated mothers. Relatedly, Black et al. (2010) reported that preschool attendance in Norway led to a significant improvement in the national exam test scores among those age 16, with the largest effects for children from low-income families. Dumas and Lefranc (2010) analysed a large-scale expansion of preschool in France and found it had positive effects on grade repetition, test scores, high school graduation, and later on adult wages. These effects were particularly large for children from a low socioeconomic background. Bingley and Westergaard-Nielsen (2012) also found that preschool attendance in Denmark was positively associated with completed schooling and earnings for those age 22-30. Preschool influence on educational attainment was larger for disadvantaged children, particularly for those with less-educated mothers.

In Uruguay, Berlinski et al. (2008) analysed a massive government-led expansion of public preschool during the late 1990s and early 2000s. Using data on children ages 7-15, the authors revealed that children who attended preschool were more likely to be enrolled in school and to have completed 0.8 additional years of schooling by age 15. The estimated effects were particularly large for disadvantaged children with less-educated parents or those living outside the capital city, Montevideo. In all, most prior work that investigated the medium- to long-term effects of ECE demonstrated significant beneficial effects, especially on educational attainment, which were largely pronounced for children from a disadvantaged background. Some differentials by outcome variables or sub-groups are not surprising because of the variety of ECE programmes and the varied populations they are applied to.

Low- and middle-income countries. There are relatively few examples of research in LMICs that examined the medium- to long-term effects of preschool attendance on later outcomes. The existing evidence largely investigated whether preschool benefits persist in terms of

[118] Note that listed countries (Norway, France, and Denmark) have a younger age range of preschool eligibility, typically between ages three and five.

students' educational attainment in secondary school or above. However, to the best of my knowledge, no study has yet assessed the long-term effects of preschool on adult outcomes, in that ECE programmes are still quite recent in most LMICs.

As for educational outcomes in early adolescence, three studies explored the associations between preschool attendance and academic achievement in the upper primary grades (age 11-12). Using the Young Lives Study in India (Andra Pradesh), Singh and Mukherjee (2018) showed that attending a private preschool rather than a government preschool was positively associated with higher math achievement at age 12. With respect to preschool entrance age, a child who entered preschool before the age of 4 obtained greater gains from preschool than those who entered preschool at the age of 5 or 6. The study in rural Uganda reported similarly positive and significant associations between preschool attendance and math achievement at age 11 (Grade 6) (Hungi & Ngware, 2018). Results indicate that attending preschool for at least two years was optimal in terms of boosting the academic achievement of students in rural Uganda. By contrast, Vandemoortele (2018) used the Young Lives Study in Ethiopia (Younger Cohort) and found that a positive association between preschool attendance and math achievement at age 8 declined by age 12. There was no longer a significant difference in math achievement based on preschool attendance for 12-year-old students.

In terms of educational outcomes in middle to late adolescence, studies put more emphasis on educational attainment such as school completion and grade retention than on academic achievement. Hazarika and Viren (2013) examined the relationships between participation in early childhood development programmes and subsequent school enrolment at ages 7-18 in rural Northern India and found significant and substantial increases in school enrolment. Specifically, participation in ECE programmes raised the average 7- to 10-year-old's probability of school enrolment by 31.5 percentage points, that of the average 11- to 14-year-old by 23.8 percentage points, and that of the average 15- to 18-year-old by 58 percentage points. These positive gains from preschool were particularly pronounced for boys and for those from households below the poverty line and less pronounced for children from households with an illiterate head (Hazarika & Viren, 2013). In Kenya and Tanzania, Bietenbeck et al. (2017) found small but statistically significant preschool benefits for students' educational attainment and cognitive development from age 7 to 16. Although students attending preschool in both countries likely started primary school later than their peers not attending preschool, they eventually caught up and even outperformed their non-preschool peers at ages 13-16 by one and a half more months of schooling. The authors suggested that lower dropout among preschool attendees underlies the gains in the number of grades completed. Similar patterns were observed for the relation between preschool attendance and cognitive skills, showing that preschool attendees outperformed their peers at ages 13-16 (0.10 *SD*, $p < 0.01$). Work by Krafft (2015) demonstrated that participation in ECE had a positive effect on educational attainment for 18- to 29-year-olds in Egypt. ECE participation led to a significant increase in educational attainment by approximately one additional year of schooling. This study identified a key pathway from ECE participation to later educational attainment that included improved school performance (e.g., higher test scores in primary), decreased grade repetition, and a higher chance of transitioning into secondary education.

Using the Young Lives Study in Ethiopia (Older Cohort), Woldehanna and Araya (2017) found that preschool attendees living in urban areas were 25.7 percentage points more likely to complete a secondary education at the proper age than their non-preschool peers. The authors suggested that the duration of the preschool exposure matters, as the preschool benefits were most pronounced for students who attended preschool for three years more than others. Furthermore, attending preschool for three years was associated with an increased probability of making the transition to higher education by age 18 (Woldehanna & Araya, 2017). While the previous study used a sample from the Older Cohort who were born in 1995, the current study used a sample from the Younger Cohort who were born in 2002. These two cohorts, which are seven years apart, may have different schooling experiences. For example, access to primary education in Ethiopia has increased significantly in recent years; the net enrolment rate for grades 1 to 4 (the first cycle of primary education) increased from 60.9 percent in 2004-2005 to 85.4 percent in 2011-2012 (MoE, 2005; 2012), while there is a decline in students' academic performance between the Younger Cohort and the Older Cohort, which signals a decline in the quality of education over this period (Woldehanna & Pankhurst, 2014). Overall, while identification is clearly an issue (i.e., omitted variables bias and endogeneity) in exploring the long-term effects of ECE (Ruhm & Waldfogel, 2012), many of the reviewed studies applied advanced identification techniques that have helped to mitigate these challenges.[119]

4.3 The Present Study

The purpose of the present study is to examine the predictive role of preschool attendance on the academic achievement and educational attainment of 15-year-old students in Ethiopia. Specifically, this study aims to assess whether the influence of preschool would *persist* or *fadeout* in adolescence during the period of transition from primary to secondary education. Using rich longitudinal data from the Young Lives Study in Ethiopia, I applied propensity score matching (PSM) to test my hypothesis that students who attended preschool at or around age five may show significantly higher levels of academic achievement and educational attainment at age 15 than their peers who never were exposed to early childhood education. I extended my analysis to investigate whether the preschool influence varied across three key dimensions: (1) the socio-demographic characteristics of child and family; (2) characteristics of the preschool experience; and (3) subsequent schooling environments. Derived from the ecological sources of influence variation, the first dimension focused on educational inequality, while the latter two focused on educational quality. I addressed three primary research questions, as follows:

1. *Preschool attendance and student outcomes:* What is the predictive role of preschool attendance on (1) students' academic achievement, as measured by receptive vocabulary, mathematics, and languages (mother tongue and English); and (2) students' educational attainment, as measured by the highest grade achieved and on-time grade progression at age 15? How do the relationships between preschool attendance and student outcomes evolve over time across ages 8, 12, and 15?

2. *Preschool attendance and student outcomes in relation to educational inequality:* Does the influence of preschool vary by gender, household wealth, father's education level,

[119] Various non-experimental methodologies were applied across the reviewed studies: (1) Difference-in-difference, Havnes and Mogstad (2012); (2) Regression discontinuity design, Black, Devereux, Loken, and Salvanes (2010); (3) Instrumental variables, Dumas and Lefranc (2012), Bingley and Westergaard-Nielsen (2012), Woldehanna and Araya (2017); (4) Sibling fixed effect, Berlinski, Galiani, and Manacorda (2008), Bietenbeck et al. (2017), Krafft, (2015); and (5) Propensity score matching, Goodman and Sianesi, (2005), Apps, Mendolia, and Walker (2013).

and student's prior achievement levels?

3. *Preschool attendance and student outcomes in relation to educational quality:* Does the influence of preschool vary by preschool characteristics, including preschool starting age, type, quality, and daily hours of participation? How are the relationships between preschool attendance and student outcomes mediated by subsequent schooling environments in primary school?

My first research question aimed to determine whether preschool attendance is predictive of students' academic achievement and educational attainment at ages 8, 12, and 15. A strength of the current analysis was the ability to track this relationship over students' educational trajectories, from early childhood to adolescence, even after accounting for their existing cognitive skills at age 5. In the second and third research questions, I aimed to identify the particular group that benefited most from enriched early learning experience and how to sustain the preschool influence through subsequent schooling experiences. The current analysis was guided by theories and prior studies in LMICs, those in Sub-Saharan Africa in particular.

The present study contributes to the knowledge base on early childhood education in several ways. First, this study added evidence to the limited literature on the medium- to long-term influence of preschool on student outcomes, particularly in the LMIC context. Such scarcity is possibly related to the lack of suitable data available until the Young Lives data were released. Specifically, I extended the findings from previous studies using Young Lives Ethiopia, which showed mixed results. Some studies using the Older Cohort sample found a positive relationship between preschool attendance and students' achievement in receptive vocabulary and math at age 8 (Woldehanna, 2016) and a higher chance of completing secondary education (Woldehanna & Araya, 2017); however, other studies which used the Younger Cohort sample found no significant relationship between preschool and students' math achievement at age 12 (Vandemoortele, 2018).[120] Focusing on the Younger Cohort sample, the present study used a more extensive and comprehensive measure of student outcomes up to age 15, including academic achievement in receptive vocabulary, math, and language, and the highest grade

[120] The Younger Cohorts (birth in 2002) data from Young Lives have some advantages over the Older Cohorts (birth in 1995) data since the Younger Cohorts data contain richer information about the preschool experience and these were collected when they were 5 years old, which is likely to mitigate recall problems.

completed through adequate grade progression. I also explored the differential influence of preschool across a variety of characteristics of the child, family, preschool, and primary school, which are critical dimensions for ensuring educational equity and quality that were not investigated in prior studies.

In addition, to mitigating differences in the demographic, geographic, and socioeconomic characteristics of students who selected into preschool *versus* those who did not, I compared a set of regression results that applied ordinary least square, kernel-based PSM, and kernel-based PSM within the community. By comparing the relative magnitude of the different estimates, I attempted to infer the mechanisms underlying students' selection to preschool and outcome models. The present study thus provides a better understanding of how the predictive role of preschool has evolved through students' educational trajectories from early childhood to adolescence in Ethiopia.

4.4 Data and Variables

The present paper used rich information from the Young Lives longitudinal study in Ethiopia. Young Lives is an international study that followed 12,000 children for 15 years in four low- and middle-income countries: Ethiopia, India (Andhra Pradesh), Peru, and Vietnam. With the aim of investigating the causes and consequences of childhood poverty, Young Lives Ethiopia collected data on 1,999 children aged 6 months to 18 months (the Younger Cohort), and 1,000 children aged 7.5 to 8.5 years (the Older Cohort). The first wave of the study was conducted in 2002 (Round 1), followed by four subsequent waves in 2006 (Round 2), 2009 (Round 3), 2013 (Round 4), and 2016 (Round 5).

The Young Lives dataset provides a wealth of information on children's trajectories across various domains, including cognitive development, socioeconomic status, demographics, and access to basic services; data were collected at the child, household, and community level. All information in the Young Lives Study was collected through face-to-face interviews with the main caregivers and children. For the present study, the longitudinal nature and richness of the Young Lives Study offer a unique opportunity to examine whether early learning inputs affect children's cognitive development and educational attainment at a later age. The current analysis used all five rounds of the Young Lives Ethiopia survey, in particular the Younger Cohort, for which information on preschool attendance was collected when the children were preschool-eligible age, which could mitigate concerns about recall bias or measurement error.[121]

Young Lives collected data from 20 sentinel sites located in five regions—Addis Ababa, Oromia, Amhara, SNNP, and Tigray—where 96 percent of the country's population lives. Considering the challenges of conducting child-focused longitudinal research in low-income countries, Young Lives adopted a multi-stage sampling strategy using a mixed approach of purposive and random sampling (Wilson et al., 2006). First, 20 sentinel sites were purposefully selected in poverty-prone areas based on food deficiency; to capture Ethiopia's diversity across regions and ethnicities, and in urban and rural areas; and to ensure the sustainability of the survey, which included selecting accessible rather than remote sites (Outes-Leon & Sanchez, 2008). Second, within the sites, about 100 households with a child aged 6-17 months old were randomly selected. In comparison to households in the 2000 Demographic and Health Survey, Young Lives households (Round 1 survey, 2002) had better access to basic services such as drinking water and electricity, except those in Addis Ababa. Meanwhile, in comparison to the 2000 Welfare Monitoring Survey, Young Lives households were poorer in terms of several assets, such as land, home, and livestock ownership (Outes-Leon & Sanchez, 2008). These findings are in line with the sampling design, which preferred poorer areas with better accessibility. Although Young Lives captured the country's diversity, it was not a nationally representative or sub-geographically (e.g., region, zone, or sentinel) representative sample, thus generalisation to the national Ethiopia population should be avoided.

4.4.1 Sample and Instrument Used for the Analysis

As with most longitudinal datasets, the Young Lives survey was not free from sample attrition and missing components—in particular for the item non-responses in dependent variables. From the original sample of 1,999 children, the attrition rate of the Younger Cohort in Young Lives Ethiopia was 5.3 percent across the five rounds (Young Lives, 2016), which is lower than any other longitudinal study.[122] Sample attrition occurred when the children were not

[121] The Older Cohort data in the Young Lives were collected after preschool-eligible age, when they were 7.5 to 8.5 years old, which may cause some recall problems. Previous study used the Older Cohort to examine the effect of preschool attendance on educational attainment in secondary school or higher education completion (Woldehanna & Araya, 2017).

[122] Actual attrition of younger cohort is 9.6 percent in Round 5; however, attrition rates do not include deaths, which account for 85 (4.3%) children. Building stable relationships between the families and field supervisors,

found because households moved and were impossible to track (14%), refused to take part in later rounds (10%), or migrated internationally (6%) (Young Lives, 2018). Out of the 1,812 students interviewed in Round 5, 1,803 students fully participated in the survey from 2002 (Round 1) to 2016 (Round 5).

In terms of missing components, the item non-responses for the explanatory variable (preschool attendance in this study) and control variables were almost none, at less than 0.04 percent. However, the item non-responses were particularly high for the outcomes of interest, which were measured by at least three test scores and the highest grades achieved in each round. It was first noted that non-response rates in any of the outcome variables varied by rounds: 5.0 percent for Round 3; 15.3 percent for Round 4; and 16.5 percent for Round 5. Further, missing test scores in one subject were not fully nested in missing values in other subjects within the same survey round. To illustrate, in Round 5, there were valid scores for 1,623 students in PPVT, 1,709 in math, 1,678 in language (English), and 1,687 in highest grade achieved, whereas students who had valid scores in all four outcome measures numbered only 1,447. Given that the present study focused on assessing the contribution of preschool sustained in the adolescence period, I decided to anchor the sample of the current analysis to Round 5. After accounting for attrition and item non-responses, Round 5 contained 1,447 students who had no missing values in the four outcome variables measured at age 15, approximately 72 percent of the initial survey sample in 2002 (1,999 respondents).

To accommodate missing data appropriately in the modeling process, I investigated the possible missing data mechanism before determining the final sample for this study. Understanding the nature of missing patterns and the implications of such incompleteness was important in order to properly use the available data and reduce potential bias to a minimum in parameter estimates. Pairwise correlation suggested that the missingness for the outcome variables was strongly correlated with child's ethnicity, family wealth quintile, region, household expenditure on education, and the education levels of father and caregiver. Table 4.1 presents an investigation of the missing data pattern of two key variables, household wealth and location. Non-response rates for outcome variables were especially high for households in

who were working since Round 1, contributed to keeping the 15-year attrition rate low in the Young Lives survey (Young Lives, 2018).

the lower wealth quintile and those living rural areas, which may indicate 'missing not occurring at random' (Little & Rubin, 2002).[123]

Table 4.1. Missing Response Pattern by Wealth Quintile and Urban-Rural Residence

Survey Round	Number of missing responses in any of outcome variables (%)	Missing response distribution by wealth quintile (%)					Missing response distribution by residence (%)	
		Poorest (Q1)	(Q2)	(Q3)	(Q4)	Richest (Q5)	Urban	Rural
Round 3	81 (5.0)	27.16	35.80	18.52	11.11	7.41	17.28	82.72
Round 4	265 (15.3)	31.50	39.94	22.10	14.86	5.80	17.66	82.34
Round 5	286 (16.5)	30.07	36.71	19.23	6.55	4.93	10.84	89.16

Source: Young Lives Dataset Round 3 to Round 5, Young Lives

To keep cases with missing data and possibly prevent any potential bias from it, one can consider multiple imputation, which has been a popular technique to facilitate statistical analysis of incomplete data (Rubin, 1976, 1996; Little & Rubin, 2014). Nevertheless, this approach may not be the best approach for the current analysis, for the following reasons: first, the multiple imputation has good properties if the data are 'missing at random' (Allison, 2002), whereas the non-response pattern of the Young Lives data could be 'missing not at random'. Second, if only the dependent variable has missing values and auxiliary variables are not identified, the use of multiple imputation should be avoided (Garson, 2015; Jakobsen et al., 2017), as is the case with Young Lives.[124] In this occurrence, no additional information will be obtained by using multiple imputation, but the standard errors may increase due to the uncertainty introduced by the multiple imputation (Garson, 2015; Jakobsen et al., 2017).

To manage item non-response I instead used listwise deletion (also known as complete case analysis), an approach that produces approximately unbiased parameter estimators, even when data are not missing at random (Little, 1992).[125] Although listwise deletion induces a loss of sample that results in larger standard errors, for regression analysis this approach is 'even more robust than the sophisticated methods [such as maximum likelihood and multiple imputation] to violation of the "missing at random" assumption' (Allison, 2002, p. 7). I conducted two

[123] In general, there are three possible missing data mechanisms, missing not at random (MNAR), missing at random (MAR), and missing completely at random (MCAR). If missing data are MNAR, the data missing mechanism is considered to be non-ignorable, while MCAR and MAR are considered to be ignorable (Little & Rubin, 2002).
[124] Auxiliary variables are the variables not included in the regression analysis but correlated with a variable with missing values and/or related to its missingness (Jakobsen et al., 2017).
[125] Provided that the data used in the study are not a nationally representative sample, complete case analysis will not affect the representativeness of the sample.

specification checks to reaffirm my choice of listwise deletion. I first ran the multiple imputation (20) using the PPVT scores in Round 4 (with the highest missing values) and found no meaningful difference in the results of interest. For example, the association between preschool attendance and receptive vocabulary at age 12 was nearly same before (β =0.52, SE = 0.07, Obs. = 1,620) and after (β =0.51, SE = 0.06, Obs. = 1,886) multiple imputation. Second, I ran the analysis after listwise deletion using the larger sample, holding at least one of the outcome variables valid rather than holding all outcomes variables valid. The coefficient estimates with the larger sample, which were varied by subject, tended to be larger than the restricted sample. Hence, performing the analysis with the non-imputed sample holding all outcome variables valid can be seen as a conservative decision for the present study.[126]

Consequently, the final sample of the present study is 1,447 for those who had valid responses across all outcome variables measured at age 15. All subsequent analysis was conducted by pooled sample (n = 1,447) and urban sample (n = 652). The latter is a more restricted group for the resident areas, considering the high prevalence of preschool participation in urban areas in 2007. When Young Lives children were five years old, attending a preschool was regarded as a luxury that was exclusively available to children from wealthy families (see more details on Section 2.4.3). The preschool attendance rate was in turn highly concentrated in the urban areas (57%) and very low in rural areas (3%). However, beyond the dichotomy between urban and rural location, various other factors, such as parental education level, household wealth, child's birth order, and sub-geographical factors, were considered strong determinants of preschool attendance (Woldehanna, 2016; Vandemoortele, 2018).[127] In order to achieve a plausible identification of preschool influence, I created the counterfactual scenario separately by the pooled and urban samples. In the next section on the model fit, I discuss the process of establishing a reliable counterfactual in each sample through propensity score matching.

[126] The results from two specification checks are available upon request.
[127] In the previous study using the Young Lives data, Woldehanna (2012, 2017) exclusively used the urban sample for the analysis. However, Vandemoortele (2018) used the pooled sample for the primary analysis since similar results were found between the pooled sample and the urban-restricted sample in the sensitivity check.

4.4.2 *Model Variables*

1) Key explanatory variable

Preschool attendance. Preschool attendance was measured by caregivers' responses when the Young Lives children were preschool-eligible age (4- to 6-year olds) between 2005 and 2007, before the expansion of public preschool. During the Round 2 household survey, primary caregivers reported whether their child had 'regularly attended any formal or informal preschool since the age of three'. I categorised students as having attended 'preschool' if they attended any of these programmes, regardless of service provider or institution. Caregivers who reported affirmatively were also asked to report the child's age at first enrolment, the type of preschool the child attended (i.e., private; community-based preschool run by an NGO, charity, or religious organization; government-funded preschool; public preschool run in part by student fees and in part by government funds), average number of days per week and hours per day the child attended, whether they paid for preschool, and caregiver's self-rating on the quality of the care and teaching at the preschool.

Table 4.2. presents the preschool characteristics used in sub-group analyses. To facilitate these analyses, preschool starting age reported in months was divided by three age groups (4, 5, and 6 or above). The type of preschool was then regrouped by private, public (government), and community-based preschool, as suggested by Orkin et al. (2012).[128] Quality of preschool care and teaching were categorised as excellent, good, and reasonably okay or bad. Average daily hours the child attended were divided into full-time (7 hours and above) and part-time (less than 7 hours), the cut-off point being seven hours per day. In all, about three-quarters of Ethiopian children with preschool experience started preschool before age six, attended a preschool run by a private organization that provided fairly good-quality service, and spent at least seven hours per day at preschool.

Some caution must be paid in interpreting the quality indicators for preschool, which is a subjective measure which relies on caregivers' responses to the question, 'In your opinion, how good is the quality of the care and teaching at this preschool?' In general, parents paying for preschool tended to believe the quality of education was high. As the evidence from high-

[128] Orkin et al. (2012) suggest that private schools and public fee-paying schools can be grouped together under 'private' preschool due to their similarities, while government preschools could be renamed 'public preschools'.

income countries suggests, parents are likely to report their child as receiving a high-quality early childhood education (NICHD & Duncan, 2003). Despite the upward tendency in parents' reporting, some gradients in the preschool quality ratings were observed in the Young Lives survey: 23.6 percent of caregivers rated the preschool as excellent, 53 percent as good, and 23.4 percent as reasonably okay or bad. Given the lack of reliability on this indicator, the analysis using preschool quality is exploratory in nature.

Table 4.2. Preschool Characteristics Where Young Lives Children Attended

Variable	Number of Students (N = 402)	Percentage (%)
Preschool starting age		
Age 4	129	32.1 %
Age 5	205	51.0 %
Age 6 or above	68	16.9 %
Preschool type		
Private	317	78.9 %
Public (Government)	24	5.9 %
Community-based	61	15.2 %
Quality of care and teaching at pre-school		
Excellent	95	23.6 %
Good	213	53.0 %
Reasonably okay or (extremely) bad	94	23.4 %
Time spent in preschool		
Full-time (>= 7 hours per day)	304	75.6 %
Part-time (< 7 hours per day)	98	24.4 %

Source: Young Lives data, Round 2 (2008), Young Lives

2) Outcomes of interest

Academic achievement. Three subjects were used to measure students' academic achievement, based on assessments of receptive vocabulary, mathematics, and languages (mother tongue for Rounds 3 and 4, English for Round 5). Table 4.3 summarizes the measures for academic achievement administered from Rounds 3 (age 8, 2009), 4 (age 12, 2013), and 5 (age 15, 2016). Students' test scores in Round 2 (age 5, 2006), as measured by two assessments—the Peabody Picture Vocabulary Test and the Cognitive Development Assessment for Quantity—were used as control variables, and for the sub-group analysis in the present study, which denotes students' cognitive skills before entering primary school.

The Peabody Picture Vocabulary Test (hereafter PPVT) is a test of receptive vocabulary recognition (Dunn & Dunn, 1997), which was administered in all survey rounds. The test consists of selecting a picture that best represents the meaning of a stimulus word the examiner presents orally (Cueto & Leon, 2012). PPVT, which has been widely used as a general measure of cognitive development, was adapted to the Ethiopian context and translated into local languages.[129] The math and language assessments were based on the Ethiopia's national curricula and international testing programmes (e.g., EGRA, TIMMS), and the test items were ordered by increasing difficulty.

Table 4.3. Assessments for Academic Achievement Administered in Rounds 3 to 5

Survey	Age	Receptive Vocabulary	Mathematics	Language
Round 3 (2009)	age 8	PPVT (204 words, ceiling applied)	Math test: 29 items in basic numeracy and math operations	Mother Tongue: 14 items in reading and listening comprehension
Round 4 (2013)	age 12	PPVT (selected 55 words)	Math test: 28 items in math operations (e.g., addition, division, square roots) and problem solving	Mother Tongue: 24 items in reading comprehension
Round 5 (2016)	age 15	PPVT (selected 55 words)	Math test: 30 items in math operations and problem solving	English: 27 items in reading comprehension

Source: Young Lives Dataset Round 3 to Round 5, Young Lives

Based on the assessments shown in Table 4.3, I used the 'percentage of correct answers' and 'standardised test scores' (*z-score*) to measure students' academic achievement. The estimates using standardised test scores should be interpreted as students' relative performance (standard deviation), where an individual's test score is compared to the average score within the sample. In that various measures were administered in each round, it is challenging to establish the cross-round comparability of students' outcomes by generating adequately equated test scores over time—that is, the application of item response theory. To illustrate, the test administration rules and items varied by each round and by subject, especially in PPVT. In the first two rounds (Rounds 2 and 3), PPVT provided test-takers with enough items (204) to set both a baseline and a ceiling, then calculated the percentage of correct answers. However, in the latter two rounds (Rounds 4 and 5), only 55 selected items were administered to students, rather than establishing a baseline or ceiling.[130] Although some anchored items were identified and equated across the survey rounds (Leon & Singh, 2017), this could not fully address the issue

[129] PPVT is also adaptable according to age from 2.5 to adulthood.
[130] The selection of PPVT items in Rounds 4 and 5 considered the similar cognitive equivalence and the level of difficulty from the previous PPVT instruments in Rounds 2 and 3 (Leon & Singh, 2017).

raised by changes in the administration rules and items, as well as the technical difficulties of equating items in six different mother tongues in Ethiopia.[131]

As for the math assessments, since the math tests in Rounds 4 and 5 added higher order numeracy skills (e.g., problem-solving) with descriptive questions, anchored items across three rounds were limited to just a few computational skills (e.g., 2+3, 15 × 9, 27÷3). Moreover, while each child took the test at his or her own pace, there were differences in the point at which the test was stopped due to the test item difficulty. The math test in Round 3, that included 29 items, was discontinued after 8 minutes, while the math tests in Round 4 (28 items) and Round 5 (30 items) were discontinued after 40 and 50 minutes, respectively.

More variations were observed in the language assessments, which reflected students' reading skill development and language of instruction policy in Ethiopia. Round 3 adopted three EGRA sub-tasks, including oral reading fluency (correct words per minute) and reading and listening comprehension tests. Round 4 conducted reading comprehension tests and simple reading tests in the mother tongue, Amharic (Ethiopia's official language), and in English. To reflect a change in language of instruction in secondary education, Round 5 administered reading comprehension tests only in English. According to the language of instruction policy in Ethiopia, mother tongue languages are used as the medium of instruction up to Grade 4, while students start to learn English as a subject from kindergarten or Grade 1 (lower primary); teachers use English as the medium of instruction in certain subjects, such as mathematics, up to Grade 8 (upper primary); and English becomes the main medium of instruction for all subjects starting in Grade 9 (secondary education). For the main analysis, I used the language assessment of reading and listening comprehension that were consistently administered in Rounds 3, 4, and 5 (I used some additional tests from Rounds 3 and 4 for the robustness check only).

Educational attainment. Educational attainment was measured by two indicators: highest grade achieved, and on-time grade progression by the time of Rounds 4 and 5. The highest grade achieved was reported by the respondents at the ages of 12 and 15. On-time grade progression measured whether a child attended the 'right' grade for her or his age by

[131] Leon and Singh (2017) provide the IRT (3PL) estimates of PPVT scores across rounds and cohorts in only three local languages in Ethiopia (Amharic, Tigrinya, and Oromifa).

progressing without any interruption, such as late school entry, grade repetition, or dropout.[132] For example, this dummy variable takes a value of 1 if a 12-year-old student is currently enrolled in Grade 6, and 0 otherwise.[133] In Round 5, when Young Lives students turned 14 or 15 years old, completing Grades 7 to 9 was regarded as on-time grade progression. Meanwhile, educational attainment in Round 3 (2009) was excluded from the outcome measures because of the ambiguity in measuring on-time grade progression between ages 7.5 and 8.5, as the compulsory age for school entry in Ethiopia is age 7.[134]

3) Model controls

The present study included a set of control variables reported by primary caregivers in Rounds 1 (2002) and 2 (2008) of the Young Lives survey, given that only variables unaffected by preschool participation should be included (Caliendo & Kopenig, 2008). Control variables were selected for inclusion in the model (matching procedure) based on prior research suggesting their potential relation with preschool attendance and/or student learning outcomes. The selected covariates can be grouped into four main categories: (1) household wealth; (2) household sociodemographic and geographic characteristics; (3) student characteristics; and (4) student's prior academic achievement.

Household wealth. The positive association between socioeconomic status and school participation and child outcomes in LMICs has been well established (Black et al., 2017; Grantham-McGregor et al., 2007; Lewin, 2009; Lopez Boo, 2016; Naudeau et al., 2011). Prior work using the Young Lives data has found that household wealth is a strong predictor of preschool enrolment in Ethiopia (Woldehanna, 2016; Vandemoortele, 2018). Before 2010, private service providers in the pre-primary sector only served middle- to high-income parents. Specifically, as the wealth index rises by one unit, a child's probability of being enrolled in

[132] Instead of using the self-reported data on grade repetition or dropout, I matched the child's age in months and the highest grade achieved to estimate on-time grade progression. In general, self-reported data on grade repetition or dropout tends to be lower than actual figure in the countries like Ethiopia where having automatic promotion practices for Grade 1 to Grade 4.

[133] In the Ethiopian education system, primary and secondary education both have two cycles. First, primary education consists of the primary first cycle (Grade 1 to Grade 4 for ages 7 to 11) and the primary second cycle (Grade 5 to Grade 8 for ages 12 to 14). Secondary education consists of the secondary first cycle (Grade 9 to Grade 10 for ages 15 and 16) and the secondary preparatory cycle (Grade 11 to Grade 12 for ages 17 and 18, TVET levels) (MoE, 2015).

[134] Depending on the month of child's birth, some entered primary school at the time of Round 3 and reported the highest grade as none.

preschool increases by 74.7 percent (Woldehanna, 2016). Consequently, I estimated models controlling for household wealth quintile based on the wealth index. The household wealth index was composed of three sub-indexes: (1) the housing quality index (e.g., main materials of walls, roof, and floor); (2) the access to service index (e.g., access to electricity, drinking water, and sanitation); and (3) the consumer durables index (e.g., possession of radio, television, and bicycle), all of which have equal weights in the estimation of wealth index.

Household socio-demographic and geographic characteristics. I included household socio-demographic and geographic characteristics that prior work has found to be associated with school attendance and child outcomes, including education levels of father and caregiver, birth order of child (first born), household size (larger than 6), private expenditure on child's education, caregiver's educational aspirations, language used between caregiver and child, urban and rural residence, and region (sub-geographic).

Selected control variables were aligned with evidence on the key determinants of preschool attendance in Ethiopia, including having a more educated parent, being a first-born child, speaking Amharic as a first language, and living in an urban area and/or Addis Ababa (Woldehanna, 2016; Vandemoortele, 2018). Substantial evidence points out that parental education level is an important predictor of children's educational outcomes (Davis-Kean, 2005; Duncan, Brooks-Gunn, & Klebanov, 1994; Haveman & Wolfe, 1995; Nagin & Tremblay, 2001, for high-income countries). In Ethiopia and Peru, highly educated and higher income parents invested more in their children, particularly at younger ages, and these differences in investment led to large gaps in learning inequality by age 8 that persisted through age 15 (Attanasio et al., 2017). The current analysis thus included estimated private expenditures on the child's education to approximate household investment in the child. More precisely, from the household consumption data in Round 2, the expenditures considered are for schooling fees, school uniforms, tuition payment, schoolbooks and stationery, transport to/from school, and cinema/entertainment. Also, the present study introduced a control variable, caregiver's aspirations for a child's educational outcomes ('Ideally, what level of formal education would you like your child to complete?'), which was measured by her or his responses in Round 2. Numerous studies have revealed that parents' educational aspirations or expectations have a considerably stronger relationship with students' academic achievement than other type of parental involvement (Hess et al., 1984; Peng & Wright, 1994, for high-income countries). In

Ethiopia, Favara (2016) found an intergenerational transmission of aspirations between parents and children, and that aspirations have strong predictive power for later educational attainment, particularly for boys.

Student characteristics. Based on evidence from prior studies, I included student characteristics that entail child age, gender, nutrition status as measured by height-for-age scores, health problems before or at age 5, and child's ethnicity.[135] In Ethiopia, being male and better nourished are positively associated with preschool attendance (Vandemoortele, 2018). Poor nutrition status has a strong link with household poverty and prevalence of illness. In fact, many children in Ethiopia remain undernourished, and a lack of good health facilities results in frequent illness and gaps in schooling, which lead to school dropout for vulnerable children (UNESCO, 2014). Lastly, poor educational participation and progression remain a challenge for certain ethnic minority groups in Ethiopia, issues that often are compounded by geographic and regional disparities (Tesfay & Malmberg, 2014).

Student prior academic achievement. Students' prior academic achievement was included, which has been associated with both early schooling and student outcomes. As supported by dynamic skill formation models (Carneiro & Heckman, 2003; Cunha & Heckman, 2007), disparities in early cognitive development between advantaged and disadvantaged students are the principal sources of divergence in their future education. Singh (2014) highlighted that, across the Young Lives countries (Ethiopia, Andra Pradesh in India, Peru, and Vietnam), disparities in academic achievement open up by the age of five and accumulate throughout the educational trajectory. In these countries, students from the richest quartile made more progress than those from the poorest quartile in mathematics during their transition from home (age 5) to primary school (age 8) (Rolleston et al., 2013). In the present study, students' prior cognitive skills before school entry were measured by their test scores at age 5 in two assessments, including PPVT and Cognitive Development Assessment for Quantity. Meanwhile, there is a possibility that some children were attending preschool at the time their skills were assessed. In that case, the measured skills could be the first manifestation of the effects of preschool attendance on learning. For inclusion in the matching model, students were divided into three

[135] The height-for-age standardised score (z-score) is a popular measure of nutrition status. For instance, a child's stunting is defined by low height-for-age, that is, below -2 standard deviations from the reference category (WHO, 2011).

ability groups (high, middle, low) according to their standardised scores on each assessment.

4.5 Empirical Strategy: Propensity Score Matching

The present study used propensity score matching, one of the quasi-experimental methods designed to evaluate treatment effects in the absence of a control group. Developed by Rosenbaum and Rubin (1983) in the early 1980s, PSM emulates a situation which experimental research achieves through randomization by modeling the treatment assignment patterns directly and then creating sub-groups which match in their likelihood of belonging to either a treatment or a control group (Guo & Fraser, 2015). PSM models have been used in numerous studies that evaluated the effect of ECE (e.g., Armecin et al., 2006; Goodman & Sianesi, 2005; McCoy, et al., 2017), thus they circumvent the problem of selectivity in preschool assignment based on various contextual factors, such as household wealth, parental education, and geographic disadvantages.

The primary goal of PSM is to match an individual who received a treatment with a comparable individual from the non-treated group who is as similar as possible in observable and pre-treatment characteristics. In the present study, this approach allowed for the identification of an adequate counterfactual that led to a comparison of children who attended preschool with children who shared similar socio-demographic characteristics but did not attend preschool. PSM assumes that the propensity score is a composite value that summarizes all observable background characteristics and that the only remaining relevant difference between the two groups is treatment—which in this study is the probability of attending preschool. According to Rubin (1997), 'the basic idea of propensity score methods is to replace the collection of confounding covariates in an observational study with one function of these covariates, called propensity score. This score is then used as if it were the only confounding covariates' (p. 461). Compared to traditional matching methods that construct an artificial comparison group across multiple characteristics, which often are vulnerable to the 'curse of dimensionality' (Reynolds & Desjardins, 2009), PSM has the advantage of matching on a single dimension (Rubin, 1997).

Through the application of PSM, the current study aimed to identify the average effect of preschool on those students who attended preschool; that is, the average effect of treatment on the treated (ATT). The ATT specifically focused on the subset of individuals actually observed to attend preschool, and thus compared their educational outcomes to an estimate of what would have happened if the same students had not attended preschool. The ATT should be distinguished from the average treatment effect (ATE), which measures the average effect for all students by the difference between treated and untreated individuals whether or not they attended preschool. Provided that preschool was not randomly assigned, the estimation of ATE may be more likely to be biased and so have fewer practical implications. The ATT is identified as follows:

$$\text{ATT} = E(Y_1 |\, T = 1) - E(Y_0 |\, T = 1) = E(Y_1 - Y_0 |\, T = 1) \qquad (1)$$

where $E(Y_1 |\, T = 1)$ represents the probability that an outcome, Y_1, will occur for those students receiving a treatment, T. In the present study, the observed outcomes Y_1 were measured by academic achievement and educational attainment at age 15 for those students who attended preschool (T). Meanwhile, $E(Y_0 |\, T = 1)$ represents the probability that these outcomes would have occurred if those same students had not received the treatment. This is a hypothetical case which assumes that those who actually attended preschool had not attended preschool as a credible counterfactual. The last part of the right-hand side captures how the ATT was estimated by the probabilities between these two plausible cases. The ATT estimates are more effective than simple parametric regression estimators for isolating preschool effects, which often fail to control for a spurious correlation between the treatment and the outcome variables (Morgan & Winship, 2007).

PSM includes two fundamental steps: the estimation of the propensity score, and the matching procedure of applying a matching algorithm using differences in the propensity score. First, I used a logistic regression model to estimate each student's propensity score. While the propensity score can be estimated by either logit or probit models, the choice between two models was not critical for the current analysis, as the two models usually yield similar results for a binary treatment case, such as preschool *versus* non-preschool (Caliendo & Kopenig, 2008).[136] Using the logit regression model, each student's propensity score was estimated as follows:

[136] I also estimated a probit model for the propensity score estimation, but the results from the probit model used a smaller number of observations under the common support area than the logit model. Hence, the logit model was used for the final estimation.

$$log\frac{P_s}{1-P_s} = \alpha + \beta_1 X_{1s} + \beta_2 X_{2s} + \cdots + \beta_\kappa X_{ks} + \epsilon_s \qquad (2)$$

where P_s is the probability of attending school, which is the estimated propensity score for student s; α and β_1 through β_κ are estimated coefficients; X_1 to X_k are covariates that include a set of observed background characteristics; and ϵ_s represents an error term that is logically distributed. Each student's propensity score was estimated once, done separately for the pooled and urban samples, and employed in all matching models in the current study.

As a second step, researchers have the choice of a matching algorithm using differences in the propensity score estimated above. There are different types of PSM (Caliendo & Kopeinig, 2005; Guo & Fraser, 2015), including (a) exact matching (i.e., pair or one-to-one matching; matches two cases with the same propensity score); (b) nearest neighbour matching (matches each treatment case with the non-treatment case that has the most similar propensity score); (c) caliper and radius matching (a type of nearest-neighbour matching that sets a limit on the distance between propensity scores to match two cases); (d) stratification and interval matching (creates intervals first, then matches treatment cases with control cases within each interval); and (e) kernel and local linear matching.

In the present study, I accomplished matching by applying kernel-based matching (Heckman, Ichimura, & Todd, 1998) using rich information from the Young Lives Study. Kernel-based matching is a non-parametric estimation approach that compares the outcome of preschool attendees to a weighted average of the outcomes of all non-attendees, based on the distance of their propensity score. The distance is measured by the difference in propensity scores between two groups, with the highest weight being placed on those with scores nearest to the particular preschool attendees. Students who are similar in their estimated propensity count more in the estimation of the treatment effect than students who are different. Kernel matching uses more information for each match and thus produces lower variance, and so it yields more precise estimates than a traditional PSM approach, such as nearest-neighbour or caliper matching (Frolich, 2004; Gasper et al., 2012).

According to Reynolds and Desjardins (2009), the ATT for matching methods is represented by

$$\text{ATT} = \frac{1}{n_1} \sum_{i \in (T=1)} \left(Y_{1i} - \sum_{j \in (T=0)} w(i,j)\, Y_{0j} \right) \qquad (3)$$

where n_1 is the number of treated cases for students who attended preschool, and $w(i,j)$ is the weight placed on each student who did not attend preschool (j) for a student who attended preschool (i). In particular, $w(i, j)\, Y_{0j}$ measures the weighted average of the outcomes for all non-preschooler cases that match to preschooler i by weighting the propensity score differentially or using different weights of $w(i, j)$. $\sum_{j \in (T=0)} w(i,j)\, Y_{0j}$ sums across all non-preschooler cases $j \in (T = 0)$ and is a key element of kernel-based matching because it implies that each preschooler case matches *all* non-preschooler cases that fall into the common support region, rather than dropping some cases that occurred in 1-to-1 or 1-to-n matching (Guo & Fraser, 2015). The weight to a kernel function is defined as

$$K(\varphi), \varphi = \frac{P_i(X) - P_j(X)}{h} = \frac{\left(\frac{P_i(X)}{1 - P_i(X)}\right) - \left(\frac{P_j(X)}{1 - P_j(X)}\right)}{h} \qquad (4)$$

in which φ represents the quality of the match. This quality of match is measured by the distance between the propensity scores of the preschooler and non-preschooler cases as a number of observations (proportion) falling into the bandwidth h. The last part of the right-hand side shows that I used the odd ratios of the propensity score for the treated and control groups (see eq. 2), which provides more robust estimates when the study sample is not representative of the overall population (Smith & Todd, 2005), as is the case with Young Lives. Using this kernel function, the weight placed on each student who did not attend preschool is

$$w(i,j) = \frac{K(\varphi)}{\sum_j K(\varphi)} \qquad (5)$$

where the kernel weight for non-preschooler cases is divided by the sum of the kernel weights, given that the matching weights $w(i,j)$ must sum to one (Reynolds & DesJardins, 2009). For the application of kernel-based PSM, researchers also have to choose the kernel function and bandwidth size (h) that consequently affect the imposition of common support (Blundell, Dearden, & Sianesi, 2005). I used the Epanechnikov kernel function, an approach that places great weight exponentially (i.e., the square of the match quality, φ) on a wider range of close matches, which in turn puts more emphasis on those comparison students with the most similar propensities (Reynolds & DesJardins, 2009). The Epanechnikov kernel function needs to be distinguished from other commonly used kernel functions that place weight based on linearity (the triangle kernel) or normal density functions (the Gaussian kernel).

Choice of bandwidth size is a 'compromise between a small variance and an unbiased estimate of the true density function' (Caliendo & Kopeinig, 2005, p. 11). In other words, wider bandwidths lower the variance, since more data are being used to construct the counterfactual, but it also lowers the match quality and increases bias. Given that there is no single rule for the bandwidth choice, I tried a range of bandwidths that yielded the best 'covariate balance' between the preschooler and non-preschooler groups, which is one of the key assumptions for the PSM (see Section 4.6.2 for details). I finally estimated models with bandwidths of 0.1-0.11; for a treated student who attended preschool, the comparison match was derived from students who did not attend preschool and whose propensity score fell within 0.05 and 0.055 on either side of the treated individual's score (Alcott, 2017). I reaffirmed my choice of bandwidths through the sensitivity check, which showed a higher sensitivity to the omitted variables than another bandwidth range (see Appendix C).

Lastly, it should be noted that correctly calculating standard errors in the PSM is a problem, for several reasons (Heckman, Ichimura et al., 1997). Through the matching procedure, the observations are no longer independent of each other, regardless of the type of matching approach applied. In other words, if there are correlations between matched pairs, standard errors are subject to bias. To correct for this bias, standard errors are adjusted by bootstrapping with 1,000 replications. All subsequent analyses were conducted using the *PSMATCH2* programme (Leuven & Sianesi, 2003) within the Stata 14.1 statistical package.

Despite the growing popularity of PSM across social science disciplines such as public health, economics, and education (Li, 2012), this approach has methodological drawbacks that should be stated explicitly. Several researchers have noted that the PSM approach has potential problems that are caused by unobserved variables, due to its high dependency on observed characteristics of the sample used in the process of matching (Michalopoulos, Bloom, & Hill, 2004). Recently, King and Nielsen (2018) argued that PSM can increase imbalance, inefficiency, model dependency, and statistical bias more than other matching methods. The main challenge stems from the fact that PSM tries to emulate a *completely randomised experiment,* which is much less flexible than a *fully blocked randomised experiment* used in the Mahalanobis Distance or other matching methods. This complete randomization attempt could lead PSM to fail to account for the large imbalance that can be eliminated by the fully blocked randomization. This occurrence is called the *PSM paradox,* which refers to the fact that 'the more balanced the data, or the more balanced it becomes by pruning some observations through matching, the more likely propensity score matching will degrade inferences' (King & Nielsen, 2018, p. 1). Although PSM carries some methodological disadvantages, it can help reduce bias more than a standard regression adjustment.

4.6 Descriptive Statistics and Model Fit

4.6.1 Descriptive Statistics

Table 4.4 provides the descriptive statistics for the variables used in my analysis, including a comparison between students who attended preschool ('Preschool' column) and those who did not ('No-Preschool' column). All estimates are presented in the pooled sample (left) and the urban sample (right). The first row shows that about 28 percent of Young Lives students reported that they attended preschool, largely in urban areas (58% among the urban sample). This average is higher than the gross enrolment ratio in pre-primary education reported by the national education statistics, which was merely 3 percent in 2007.[137] This large discrepancy in preschool participation may stem from the fact that there were issues of measurement in the national education statistics, which had difficulty capturing privately run or community-based kindergartens (MoE, 2008). On the other hand, the Young Lives sample tends to have better access to services (e.g., health, education, and social protection) than the nationally representative samples in the Welfare Monitoring Survey (Outes & Leon, 2008).

Most of covariates differed considerably between the preschool and non-preschool groups, according to the results of the *t-test* shown in the last column of each sample (Table 4.4). In particular, students who attended preschool were far more advantaged than those who did not. Among the pooled sample, for example, the household wealth gradients in preschool

[137] 2005-2007 is the year when the Young Lives children were of preschool-eligible age.

attendance were apparent: about half of students from the most affluent families had attended preschool, whereas no children from the poorest families had attended preschool. Consistent with findings from prior work (Woldehanna, 2016; Vandemoortele, 2018), most of the covariates on household characteristics showed statistically significant differences that favoured preschoolers, including richer households, parents' higher education levels, being a first-born child, living in a smaller household (fewer than 6), higher household spending on a child's education, parents' higher educational aspirations, different language use between home and school, and living in an urban area and/or Addis Ababa.[138]

As for student characteristics, while there was no significant difference by age, gender, or health problems in early childhood, students with better nutrition status were more likely to attend preschool: about half of the children in the high height-for-age group attended preschool, in contrast to only one-fifth of children in the low height-for-age group. Wide variations across children's ethnicities show the significant differences in preschool attendance between and within ethnicities. In terms of academic achievement at age 5, students in the high-performing group had attended preschool: on average, 63 percent of children with high receptive vocabulary skills reported that they had attended preschool, while 47 percent of children with low receptive vocabulary skills reported that they had not.

At the bottom of the table, I show the descriptive statistics for the outcome variables on students' academic achievement and educational attainment by different rounds. Before any adjustment to account for child and family characteristics, there were stark learning disparities between preschoolers and non-preschoolers, all of which were statistically significant (p <0.01). In all, the significant dissimilarities between preschoolers and non-preschoolers reaffirm that preschool attendance was not independent of a wide array of child and family characteristics. This descriptive picture implies that, without an appropriate analytical strategy, the estimates will be biased, due to several non-random sources of selection into preschool.

[138] On the variable of same language use between home and school, non-preschoolers (92%) were more likely to use same language between home and school than preschoolers (68%).

Table 4.4. Descriptive Statistics

Variable	POOLED SAMPLE							URBAN SAMPLE						
	Average		Preschool		No-Preschool		*t-test*	Average		Preschool		No-Preschool		*t-test*
	Mean	*SD*	Mean	*SD*	Mean	*SD*	*Diff.*	*Diff.*	SD	Mean	SD	Mean	SD	*Diff.*
Explanatory variable														
Preschool Attendance	0.28	0.45	-	-	-	-	-	0.58	0.49	-	-	-	-	-
Covariate – Household wealth														
Wealth quintile														
Quintile 1 (Poorest)	0.20	0.40	0.00	0.07	0.28	0.45	0.27***	0.21	0.40	0.12	0.32	0.33	0.47	0.21***
Quintile 2	0.20	0.40	0.02	0.16	0.27	0.44	0.24***	0.19	0.40	0.16	0.37	0.24	0.43	0.07*
Quintile 3	0.20	0.40	0.15	0.36	0.22	0.41	0.07**	0.20	0.40	0.24	0.43	0.15	0.36	-0.09**
Quintile 4	0.20	0.40	0.35	0.48	0.14	0.35	-0.21***	0.20	0.40	0.22	0.41	0.18	0.38	-0.04
Quintile 5 (Richest)	0.20	0.40	0.47	0.50	0.09	0.29	-0.38***	0.19	0.40	0.26	0.44	0.10	0.30	-0.17***
Covariate – Household characteristics														
Father's highest education level														
No education	0.27	0.45	0.12	0.33	0.33	0.47	0.21***	0.20	0.40	0.12	0.32	0.31	0.46	0.19***
Primary education	0.53	0.50	0.39	0.49	0.58	0.49	0.19***	0.43	0.50	0.38	0.49	0.50	0.50	0.12**
Secondary education and above	0.20	0.40	0.49	0.50	0.09	0.28	-0.40***	0.37	0.48	0.50	0.50	0.19	0.39	-0.32***
Caregiver's highest education level														
No education	0.50	0.50	0.17	0.38	0.62	0.48	0.45***	0.27	0.45	0.16	0.37	0.43	0.50	0.27***
Primary education	0.39	0.49	0.52	0.50	0.34	0.48	-0.18***	0.50	0.50	0.52	0.50	0.47	0.50	-0.05
Secondary education and above	0.11	0.31	0.31	0.46	0.03	0.18	-0.28***	0.23	0.42	0.32	0.47	0.10	0.30	-0.22***
First born in household	0.26	0.44	0.44	0.50	0.20	0.40	-0.24***	0.33	0.47	0.44	0.50	0.19	0.39	-0.25***
Household size (larger than 6)	0.37	0.48	0.27	0.45	0.40	0.49	0.13***	0.32	0.47	0.28	0.45	0.37	0.48	0.09*
Private Spending Level on Education														
High	0.33	0.47	0.57	0.50	0.24	0.43	-0.33***	0.33	0.47	0.43	0.50	0.20	0.40	-0.23***
Middle	0.33	0.47	0.27	0.45	0.36	0.48	0.09**	0.33	0.47	0.33	0.47	0.34	0.47	0.01
Low	0.33	0.47	0.16	0.36	0.40	0.49	0.25***	0.33	0.47	0.24	0.43	0.46	0.50	0.22***
Parental aspiration toward education	0.70	0.46	0.92	0.28	0.61	0.49	-0.30***	0.86	0.35	0.93	0.25	0.76	0.43	-0.17***
Language between home and school	0.86	0.35	0.68	0.47	0.92	0.26	0.24***	0.75	0.44	0.66	0.47	0.86	0.35	0.19***
Living in urban	0.45	0.50	0.94	0.23	0.26	0.44	-0.68***	-	-	-	-	-	-	-
Region														
Living in Addis Ababa	0.16	0.37	0.56	0.50	0.01	0.11	-0.55***	0.36	0.48	0.59	0.49	0.04	0.21	-0.55***
Living in Tigray	0.23	0.42	0.02	0.14	0.32	0.47	0.30***	0.14	0.35	0.02	0.14	0.30	0.46	0.28***
Living in Amhara	0.23	0.42	0.08	0.27	0.28	0.45	0.20***	0.12	0.33	0.08	0.28	0.17	0.38	0.09***
Living in Oromia	0.20	0.40	0.11	0.32	0.24	0.43	0.13***	0.12	0.33	0.08	0.27	0.19	0.39	0.11***
Living in SNNP	0.17	0.37	0.22	0.41	0.15	0.35	-0.07**	0.24	0.43	0.22	0.41	0.28	0.45	0.06
Covariate – Child characteristics														
Age 14	0.36	0.48	0.32	0.47	0.38	0.49	0.06*	0.35	0.48	0.32	0.47	0.38	0.49	0.05
Age 15	0.64	0.48	0.68	0.47	0.62	0.49	-0.06*	0.65	0.48	0.68	0.47	0.62	0.49	-0.05
Female	0.47	0.50	0.44	0.50	0.49	0.50	0.04	0.48	0.50	0.45	0.50	0.52	0.50	0.07
Height-for-age z-score at age 5														
High	0.33	0.47	0.48	0.50	0.28	0.45	-0.20***	0.33	0.47	0.38	0.49	0.26	0.44	-0.12**

Middle	0.33	0.47	0.31	0.46	0.34	0.47	0.03	0.33	0.47	0.36	0.48	0.30	0.46	-0.06
Low	0.34	0.47	0.21	0.41	0.39	0.49	0.17***	0.33	0.47	0.26	0.44	0.44	0.50	0.19***
Child has health problem at age 5	0.09	0.29	0.08	0.27	0.09	0.29	0.01	0.08	0.27	0.08	0.27	0.08	0.27	0.00
Child's ethnicity														
Ethnicity 1 Others	0.12	0.32	0.17	0.37	0.10	0.30	-0.07***	0.11	0.32	0.17	0.37	0.04	0.20	-0.13***
Ethnicity 2: SNNP	0.08	0.27	0.10	0.31	0.07	0.26	-0.03*	0.16	0.37	0.10	0.30	0.24	0.43	0.14***
Ethnicity 3: Oromo	0.22	0.42	0.29	0.46	0.20	0.40	-0.10***	0.24	0.43	0.28	0.45	0.20	0.40	-0.08*
Ethnicity 4: Tigrian	0.26	0.44	0.10	0.30	0.32	0.47	0.22***	0.19	0.39	0.10	0.30	0.31	0.46	0.21***
Ethnicity 5: Amhara	0.32	0.47	0.34	0.47	0.32	0.47	-0.02	0.29	0.46	0.35	0.48	0.22	0.41	-0.13***
Covariate – Child prior achievement														
PPVT at age 5														
High	0.31	0.46	0.63	0.48	0.19	0.39	-0.44***	0.31	0.46	0.41	0.49	0.17	0.37	-0.24***
Middle	0.32	0.47	0.26	0.44	0.34	0.47	0.08**	0.33	0.47	0.37	0.48	0.29	0.45	-0.08*
Low	0.37	0.48	0.11	0.31	0.47	0.50	0.36***	0.36	0.48	0.22	0.42	0.55	0.50	0.32***
CDA-Q(Math) at age 5														
High	0.26	0.44	0.52	0.50	0.16	0.37	-0.36***	0.26	0.44	0.38	0.49	0.09	0.28	-0.29***
Middle	0.38	0.49	0.34	0.48	0.40	0.49	0.06*	0.30	0.46	0.30	0.46	0.30	0.46	0.00
Low	0.35	0.48	0.14	0.35	0.44	0.50	0.30***	0.45	0.50	0.33	0.47	0.62	0.49	0.29***
Outcome variables														
Students' academic achievement	*(% of correct ans.)*				(z-score)			*(% of correct ans.)*				(z-score)		
PPVT scores (Round 3)	0	1	0.86	0.94	-0.33	0.76	-1.19***	0	1	0.39	0.90	-0.53	0.87	-0.92***
PPVT scores (Round 4)	*71.01*	*15.30*	0.82	0.54	-0.32	0.96	-1.14***	*80.0*	*10.63*	0.36	0.75	-0.54	1.08	-0.90***
PPVT scores (Round 5)	*76.88*	*13.94*	0.82	0.51	-0.32	0.96	-1.14***	*85.55*	*9.04*	0.37	0.71	-0.51	1.11	-0.88***
Math scores (Round 3)	*24.85*	*18.64*	0.84	0.92	-0.33	0.82	-1.17***	*35.87*	*18.55*	0.30	0.91	-0.42	0.97	-0.72***
Math scores (Round 4)	*39.33*	*21.11*	0.65	0.91	-0.27	0.91	-0.92***	*50.04*	*20.10*	0.19	0.95	-0.29	1.00	-0.48***
Math scores (Round 5)	*31.64*	*14.65*	0.71	1.06	-0.27	0.83	-0.98***	*38.42*	*14.81*	0.29	1.04	-0.41	0.77	-0.70***
MT scores (Round 3)	*39.14*	*22.20*	0.55	0.98	-0.21	0.93	-0.76***	*48.71*	*22.02*	0.18	0.98	-0.25	0.98	-0.43***
MT scores (Round 4)	*29.15*	*15.13*	0.62	0.88	-0.26	0.93	-0.88***	*36.28*	*14.00*	0.21	0.94	-0.33	1.01	-0.54***
English scores (Round 5)	*54.46*	*20.57*	0.72	0.83	-0.28	0.92	-0.99***	*64.06*	*18.92*	0.35	0.85	-0.48	0.99	-0.83***
Students' educational attainment														
Highest grade achieved (Round 4)	4.80	1.57	5.50	1.11	4.51	1.63	-0.99***	5.30	1.24	5.55	1.07	4.94	1.37	-0.61***
Highest grade achieved (Round 5)	6.39	1.85	7.26	1.24	6.05	1.93	-1.21***	7.03	1.44	7.31	1.20	6.63	1.64	-0.68***
On-time grade progression (Round 4, %)	0.51	0.50	0.67	0.47	0.45	0.50	-0.22***	0.64	0.48	0.68	0.47	0.58	0.49	-0.09*
On-time grade progression (Round 5, %)	0.55	0.50	0.78	0.42	0.46	0.50	-0.32***	0.69	0.46	0.79	0.41	0.56	0.50	-0.23***

Note: (1) On students' test scores, average test scores (*in Italic*) are % of correct answer, while test scores by pre and non-preschool group and the *t*-test results are standardised scores (z-score); (2) MT stands for mother tongue. (3) Ethnicity in SNNP includes Hadiya, Sidama, Wolayta; and others includes Agew, Gurage, Kambata.
*** p<0.01, ** p<0.05, *p<0.1
Source: Young Lives Dataset Round 2 to Round 5, Young Lives

4.6.2 Propensity Score Matching: Model Fit

PSM may help reduce the potential for confounding bias, but it is important to examine the validity of the matching model before reporting ATT estimates. If matching successfully eliminates initial differences between preschool and non-preschool groups, this approach allows the tentative conclusion that the learning divergence between students is more a result of attending a preschool than previously existing differences. To establish a credible counterfactual, PSM analyses must satisfy three assumptions: conditional independence, common support, and covariate balance (Reynolds & Desjardins, 2009).

Conditional independence. The key assumption of the PSM holds that, conditional on the observed covariates used in the model, the potential outcomes are independent of whether a student attended preschool (Rosenbaum & Rubin, 1983). Adherence to this conditional independence assumption implies that there are no unobservable characteristics that affect selection into either preschool or non-preschool groups and students' potential outcomes. This is apparently a strong assumption; thus, it must be justified by the quality of the data used by researchers. Blundell et al. (2005) emphasised that the plausibility of such an underlying assumption should take into account the 'informational richness of the available dataset in relation to a detailed understanding of the institutional set-up by which selection into treatment takes place' (p. 486-487). Therefore, the more extensive and accurate the data are, the easier it is to justify the conditional independence assumption in the matching procedure. Due to the fact that the matching estimates are highly dependent on the selected observables within the available dataset, this approach has often faced criticism; however, the same crucial assumption applies to both an ordinary least square (OLS) and a matching approach, and the latter has the advantage of being implemented in a more flexible way (e.g., not imposing linearity) (Blundell et al., 2005).

The Young Lives dataset is a suitable choice for overcoming this challenge, given that it provides uniquely rich information about students, households, and communities in Ethiopia from a child's birth to adolescence. To accurately select the covariates, I relied on past theory and literature that inform preschool access in LMICs, particularly in Sub-Saharan Africa. In fact, the Young Lives Ethiopia data have been used to identify the key determinants of preschool attendance and students' academic achievement in previous studies (Woldehanna, 2016; Vandemoortele, 2018). Also, as shown in Table 4.4, most of the variables selected for the propensity score estimation were significantly different between the preschool and non-preschool groups, thus justifying my selection of variables used in the current analysis.

Further, the information on the institutional settings is an important factor in terms of satisfying the conditional independence assumption. In particular, data on the communities the households belong to could contribute to improving the propensity score estimations, otherwise the community factors will remain as potentially confounding variables.[139] To mitigate potential sources of bias from the community, I applied a kernel-based matching *within* each sentinel site (equivalent to the enumeration area of the survey) as a robustness check analysis.[140] This approach, which compares the outcomes of preschoolers and non-preschoolers within their specific geographical community, reduces the potential for confounding bias attributable to unobserved cross-community differences, including preschool availability (McCoy et al., 2017). This is particularly germane to Ethiopia's decentralised administrative structure, where communities play a pivotal role in the delivery of basic services such as health, education, and social protection programmes (Khan et al., 2014).[141]

Another useful source of information is the home learning environment. The importance of the home learning environment on child development is well established, albeit mostly in high-income countries (see more details in Bradley & Caldwell, 1995; LeFevre et al., 2009; Melhuish et al., 2008; Sylva et al., 2004). Although there is no conventional measure available for the home learning environment (e.g., the number of books at home, reading to the child, or home play activities), this is partially mitigated in the current analysis by the inclusion of two measures used in the Young Lives survey: (1) household's private expenditure allocated to child's education (e.g., purchase of children's books and school uniforms), and (2) parental aspirations for their children's education.[142] This is probably beneficial, since these two

[139] The 'community questionnaire' in the Young Lives Study collected data on educational institutions in the community (e.g., number of pre-primary institutions, primary schools); however, due to the incompleteness and inconsistency of reporting, the current analysis did not account for community-level information in the survey.
[140] In Young Lives, the 20 sentinels and 24 communities were selected purposively in 2001 (21 sentinels from Round 2 considering migrated population), and 2 or 3 communities were nested in one sentinel. I used the sentinel (enumeration area) to allow me to run fixed effect with sufficient observation within the unit.
[141] The previous study (Woldehanna & Gebremedhin, 2012) used community variables as an 'instrument' to examine the effect of preschool on receptive vocabulary skills at ages 5 and 8.
[142] These two indicators were collected in the Young Lives Round 2 when children were five years old.

indicators have the potential to serve as a proxy for some variation in home learning environments.

Nonetheless, holding the conditional independence assumption remains a challenge, since unobserved variables may still exist in relation to the variables of interest. To estimate the extent to which such unobservables may bias my ATT estimates, I tested a model's sensitivity following Rosenbaum's (2002) and Mantel-Haenszel's sensitivity analyses (Mantel & Haenszel, 1959). The details of these analyses are described in Appendix C. Overall, the results of this sensitivity analysis suggested that confounding variables were unlikely to weaken the models using students' outcomes in PPVT and educational attainment, but the models using math and language outcomes were relatively highly sensitive to hidden bias arising from unobserved variables that simultaneously affected assignment to preschool and the outcome variable.

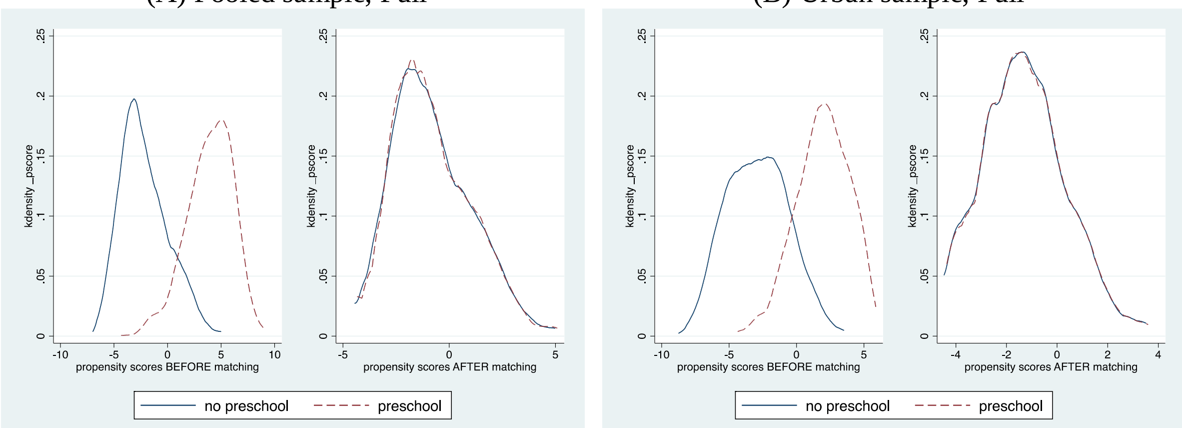
Common support. Under the common support assumption, the PSM is only appropriate when comparability of the treated and untreated groups is established by a sufficient overlap between the two along the propensity score distribution (Blundell et al., 2005; Reynolds & DesJardins, 2009). The inspection of common support can be conducted by using the test of whether comparable observations from preschool and non-preschool groups are available for the whole sample, or only in some parts of the propensity score distributions.

Figure 4.1 illustrates the common support regions, focusing on the extent to which preschool attendees and non-attendees have comparable observed probabilities of attending preschool, before and after PSM. Noticeably, the distributions of the propensity score before matching were very different between preschool attendees and non-attendees. This in turn led to the projection of a very limited area of common support, which does not exist at the tail of the distributions. It may be unsurprising, given the dissimilar features of the two groups observed for various child and household characteristics, even among children living in urban areas. To address this limited overlap in common support area, I performed a trimming procedure to delete observations at the lower and upper tails of the estimated score, as suggested by Heckman et al. (1998). By setting the value at 0.05, the matching trims participants whose estimated propensity scores are below .05 and above 0.95.[143] As explained previously, the choice of the kernel function and bandwidth size also affects the common support area; I carefully chose these two based on the covariate balance and model variance in the present study.

After the kernel-based matching, the projection presented a great deal of overlap between the two groups. However, due to the initial stark differences, the imposition of common support caused a large loss of the sample—11.4 percent and 27.5 percent, respectively, of the pooled and urban samples dropped from the common support area. In particular, the proportion of preschool attendees in the area of common support was low, whereas no one dropped from the non-preschool group within the common support area. As discussed by Blundell et al. (2005), in kernel-based matching it is typical that 'those treated whose propensity score is larger than the largest propensity score in the non-treated pool are left unmatched' (p. 486). Despite the loss of a treated sample that may affect the external validity of this study, the PSM created the optimal counterfactual—for instance, only two covariates are different at the 0.05 level after matching among the pooled sample.

Figure 4.1. Common Support Area for Preschool Attendees and Non-Attendees

(A) Pooled sample, Full (B) Urban sample, Full



Source: Young Lives Dataset Round 2 to Round 5, Young Lives

Covariate balance. The covariate balance assumption requires that the mean and standard deviation for each of the covariates do not differ significantly between the treated and matched untreated groups. As there is no consensus on what criteria should be used to determine

[143] According to (Crump et al., 2009), a rule of thumb for trimming is to discard all observations with estimated propensity scores outside the range between 0.1 to 0.9.

significant differences between covariates (Reynolds & DesJardins, 2009), I applied three commonly used methods to check the balance in observed characteristics between the matched groups.

First, I used a two-sample *t-test* to see the difference in means across the groups for each of the covariates. For any given covariate, balance is regarded as achieved when there is no difference between the treated and matched untreated groups at the .05 significance level. Table 4.5 presents the average of model covariates by (1) treated (mean) and all untreated sample (difference, *t-test p* value) before matching; and (2) treated (mean) and matched untreated sample (difference, *t-test p* value) after matching. Before matching among the pooled sample, 38 of the 42 covariates were significantly different at the .05 level. In contrast, only 2 of the 42 covariates were significantly different after matching, showing that the analytical model satisfies the covariate balance assumption. For the urban sample, after matching, 5 of the 42 covariates still remained significantly different between the two groups.

Second, I tested standardised differences and associated percentage bias by matched and unmatched groups (Caliendo & Kopeinig, 2008; Smith & Todd, 2005). Figure 4.2 is a visual presentation of these test results, showing that the matching procedure contributes to the convergence of associated percentage bias to zero in each of the covariates. There were no standardised differences greater than 20 percent, which is the threshold set by Rosenbaum and Rubin (1985). Similarly to the *t-test* results, the pooled sample showed better balance in terms of standardised percentage bias (closer to zero) than the urban sample.

Lastly, by using the Stata command *pscore,* as developed by Becker and Ichino (2002), I estimated the propensity score first by logit regression, then tested whether the balancing property held within the identified blocks. This process enabled me to check whether the mean propensity score was not different between the treated and control in each covariate block. In the present study, the balancing property with a full set of covariates was successfully satisfied in eight blocks in the pooled sample and five blocks in the urban sample.

Table 4.5. Comparison of Treated, All Untreated, and Matched Untreated Groups

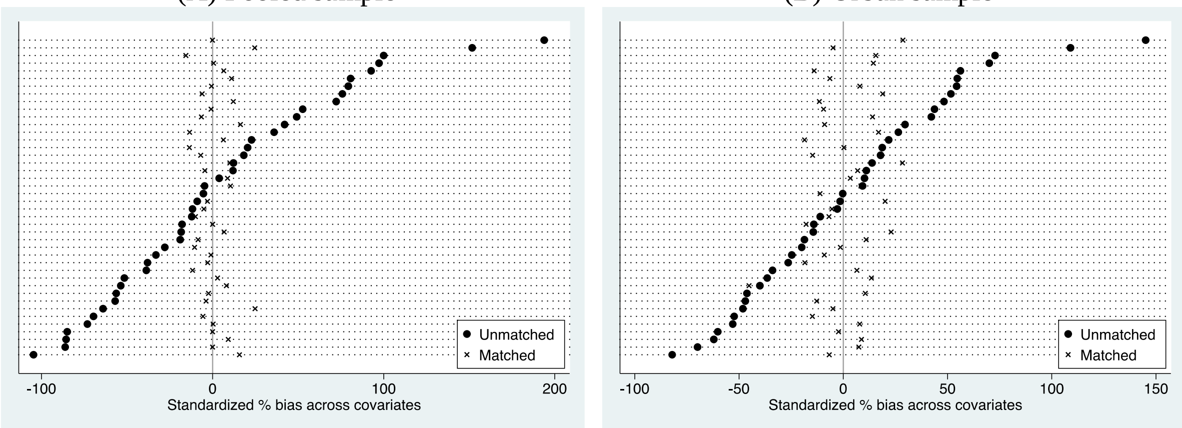
Variable	Pooled				Urban			
	Before matching		After matching		Before matching		After matching	
	Treated	All untreated	Treated	Matched untreated	Treated	All untreated	Treated	Matched untreated
	Mean	*Diff.*	Mean	*Diff.*	Mean	*Diff.*	Mean	*Diff.*
Covariate – Household wealth								
Wealth quintile								
Quintile 1 (Poorest)	0.00	-0.27**	0.01	0.00	0.12	-0.21***	0.17	0.03
Quintile 2	0.02	-0.24**	0.04	0.00	0.16	-0.07**	0.20	0.04
Quintile 3	0.15	-0.07**	0.18	0.00	0.24	0.09***	0.18	-0.07
Quintile 4	0.35	0.21**	0.33	-0.03	0.22	0.04	0.25	0.03
Quintile 5 (Richest)	0.47	0.38**	0.44	0.03	0.26	0.16***	0.21	-0.04
Covariate – Household characteristics								
Father's highest education level								
No education	0.12	-0.21**	0.16	0.01	0.12	-0.19***	0.15	0.02
Primary education	0.39	-0.19**	0.45	-0.01	0.38	-0.12***	0.45	-0.04
Secondary and above	0.49	0.40**	0.39	0.00	0.50	0.31***	0.41	0.06
Caregiver's highest education level								
No education	0.17	-0.45**	0.22	0.07**	0.16	-0.27***	0.22	0.04
Primary education	0.52	0.18**	0.55	-0.07	0.52	0.05	0.53	0.02
Secondary and above	0.31	0.28**	0.23	0.00	0.32	0.22***	0.25	-0.05
First born	0.44	0.24**	0.37	0.00	0.44	0.24***	0.35	-0.03
Household size (> 6)	0.27	-0.13**	0.31	-0.05	0.28	-0.09**	0.31	-0.01
Private Spending Level on Education								
High	0.57	0.33**	0.48	0.06	0.43	0.23***	0.31	0.09
Middle	0.27	- 0.09**	0.32	-0.04	0.33	-0.01	0.40	-0.03
Low	0.16	-0.25**	0.20	-0.02	0.24	-0.22***	0.30	-0.06
Parental aspiration	0.92	0.30**	0.89	-0.02	0.93	0.17***	0.92	-0.04
Same language btw.	0.68	-0.24**	0.68	0.09**	0.66	-0.19***	0.62	0.04
Living in urban	0.94	0.68**	0.90	-0.01	-	-	-	-
Region								
Living in Addis	0.56	0.55**	0.33	0.05	0.59	0.55***	0.39	0.11***
Living in Tigray	0.02	-0.30**	0.03	0.00	0.02	-0.28***	0.04	-0.02
Living in Amhara	0.08	-0.20**	0.14	0.03	0.08	-0.09***	0.15	-0.06
Living in Oromia	0.11	-0.12**	0.19	0.00	0.08	-0.11***	0.15	0.02
Living in SNNP	0.22	0.07**	0.31	-0.03	0.22	-0.06*	0.29	0.10**
Covariate – Child characteristics								
Age 14	0.32	-0.06*	0.33	-0.05	0.32	-0.05	0.36	-0.03
Age 15	0.68	0.06*	0.67	0.05	0.68	0.05	0.65	0.03
Female	0.44	-0.04	0.44	-0.01	0.45	-0.07*	0.42	-0.09
Height-for-age z-score at age 5								
High	0.48	0.20**	0.45	0.08	0.38	0.12***	0.37	0.08
Middle	0.31	-0.03	0.30	-0.02	0.36	0.06*	0.35	0.13***
Low	0.21	-0.17**	0.25	-0.05	0.26	-0.19***	0.29	-0.21***
Health prob. at age 5	0.08	-0.01	0.08	0.03	0.08	0.00	0.08	0.05**
Child's ethnicity								
Ethnicity 1: Others	0.17	0.07**	0.13	-0.05	0.17	0.13***	0.14	0.04
Ethnicity 2: SNNP	0.10	0.03**	0.16	-0.01	0.10	-0.14***	0.17	0.05
Ethnicity 3: Oromo	0.29	0.10**	0.30	0.03	0.28	0.08**	0.29	0.00
Ethnicity 4: Tigrian	0.10	-0.22**	0.07	-0.01	0.10	-0.20***	0.31	-0.06
Ethnicity 5: Amhara	0.34	0.02	0.34	0.04	0.35	0.13***	0.33	-0.04
Covariate – Child prior achievement								
PPVT at age 5								

High	0.63	0.44**	0.52	-0.07	0.41	0.24***	0.28	0.03
Middle	0.26	-0.08**	0.30	0.03	0.37	0.08**	0.38	-0.07
Low	0.11	-0.36**	0.18	0.04	0.22	-0.32***	0.35	0.03
CDA-Q(Math) at age 5								
High	0.52	0.35**	0.38	0.05	0.38	0.29***	0.23	0.06
Middle	0.34	-0.06*	0.41	-0.02	0.30	0.00	0.33	-0.05
Low	0.14	-0.30**	0.21	-0.02	0.33	-0.29***	0.45	-0.01

Notes: * indicate significant difference from treatment group:
*** p<0.01, ** p<0.05, *p<0.1
Source: Young Lives Dataset Round 2 to Round 5, Young Lives

Figure 4.2. Covariate Balance before and after Matching

(A) Pooled sample (B) Urban sample



Source: Young Lives Dataset Round 2 to Round 5, Young Lives

4.7 Results

4.7.1 Research Question 1: Preschool Attendance and Student Outcomes at Age 15

From Table 4.6 to Table 4.8, I illustrate the results of four sets of regression analyses used in the present study. In Model 1, an unadjusted OLS regression was estimated to show average differences in student outcomes between students attending preschool and students not attending preschool. Model 2 presented estimates of an OLS specification which controlled for a full set of covariates, which were included in Models 3 and 4.

In Model 3, estimates applying the kernel-based propensity score matching (Heckman, Ichimura, & Todd, 1998) were presented to account for observable sources of non-random selection into preschool. Through the matching process, a set of models was fit in to construct the credible counterfactual of preschool and non-preschool students, based on an extensive list of child, household, and community characteristics. Using the weights generated by the kernel-based matching, I present the 'difference in means' between treated (preschool) and control (non-preschool) groups for the matched sample.[144] It should be noted that ATT estimates in Model 3 had to be redefined as the mean difference for those falling within the common support region. Standard errors were calculated using bootstrap, with 1,000 replications. I consider this to be my primary analytic approach for the present study.

In Model 4, I replicated the kernel matching approach within each sentinel site of the Young Lives Study, comparing preschool attendees' outcomes to the weighted average of non-attendees within their specific geographic community. By comparing students within a given sentinel site, this approach reduced the potential for confounding bias caused by unobserved between-community differences. Meanwhile, Model 4 reduced the sample size because it required the within-subject variability in a given sentinel site to have at least one preschool attendee and one non-attendee. To illustrate, when this model was applied, 415 sample students from six sentinel sites were dropped, as they did not have either a preschool attendee or a non-attendee. In this regard, I consider Model 4 a robustness check to my primary analytical approach in Model 3.

Preschool and academic achievement. Table 4.6 presents the results of the relationship between preschool attendance and students' academic achievement for the pooled sample. In Model 1, the unadjusted OLS estimates of preschool attendance were largest for PPVT at age 8 (1.19 *SD*) and math at age 8 (1.17 *SD*), and smallest for mother tongue (MT) at age 8 (0.76 *SD*) and age 12 (0.88 *SD*), all of which were statistically significant (p <0.01). In Model 2, after introducing a set of covariates, the coefficients of preschool attendance were substantially reduced. The largest and significant associations were observed for PPVT at age 8 (0.43 *SD*) and math at age 8 (0.38 *SD*), whereas the associations became no longer significant for math at age 12 (0.02 *SD*) and age 15 (0.04 *SD*), and for MT at age 8 (0.15 *SD*) and age 12 (0.12 *SD*). The positive associations between preschool and students' performance only persisted up to age 15 in PPVT and English tests, with the coefficients of 0.24 *SD* for both (p <0.01), while the associations faded out on students' performance in math and MT tests from age 12 onwards.

[144] Same parameters can be estimated by applying the weights generated by the kernel matching to an unadjusted regression equation (Model 1). For the sensitivity check of the estimates obtained by 'difference-in-means' in the current analysis, I also conducted 'regression-adjusted matched estimation', which runs a regression of outcome on treatment indicator and confounding covariates (Model 2) applying the weights generated by the kernel matching. The regression-adjusted regression drew a similar result with difference-in-means (The results are available upon request).

Consistent with the OLS estimates in Model 2, Model 3, with kernel-based matching estimates (the primary approach of the present study), showed that preschool attendance was associated with significantly better performance in PPVT at ages 8, 12, and 15 and in English at age 15, yet the predictive role of preschool dissipated in math and MT performance by age 12. As for PPVT scores, the ATT estimates of preschool attendees *versus* non-attendees were attenuated from age 8 (0.36 *SD*) to age 12 (0.21 *SD*) and age 15 (0.19 *SD*), but all remained statistically significant at $p < 0.01$. Conversely, the ATT estimates in math steeply decreased from age 8 (0.51 *SD*) to age 12 (0.09 *SD*) and age 15 (0.16 *SD*) and were no longer statistically significant. Similarly, in the language assessments, the associations between preschool and MT test scores diminished from age 8 (0.29 *SD*) to age 12 (0 *SD*). One exception was the predictive role of preschool in significantly better performance in English at age 15 (0.33 *SD*, $p < 0.01$). While there is MT-centred instruction in Ethiopia during the lower primary grades, students are exposed to English from kindergarten or Grade 1. Thus, this could be seen as accumulated language skills in English, which were measured at age 15 for the first time.[145] Model 4, with the kernel-based matching estimates within sentinel sites (robustness check), showed similar patterns with Model 3. Results were only slightly different with respect to the observed magnitude and significance of the associations relative to the estimates in Model 3, except English performance at age 15.

Given that estimates from the OLS model tend to be biased upwards, the kernel matching estimates in this study (Model 3) were more conservative than the OLS estimates with full covariates (Model 2) for students' performance in PPVT. The reverse patterns were observed in math and language achievement, as the kernel matching estimates were larger than the OLS estimates. One partial explanation could be the low sensitivity of the ATT estimates to the confounding variables, which were lower in math test scores ($\Gamma = 1.1$) and mother tongue scores ($\Gamma = 1.0$) than other test scores in PPVT ($\Gamma = 2.15$) (see Appendix C for more details). On the other hand, as noted earlier, some of sample in the treated group was dropped during the matching procedure due to the lack of common support, which may change the composition of matched pairs and so lead to different estimates.

[145] Young Lives introduced the English language assessment in Round 5 (2016) and school survey 2016-2017. It reflects the growing emphasis on English learning in Ethiopia as a 'transferable skill' to continuing education, labour market opportunities, and social mobility (Graddol, 2010). Round 4 administered only a simple reading test of English skill (e.g., read a word or a sentence).

Table 4.6. Relation between Preschool and Academic Achievement, ATT (Pooled Sample)

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
Subject		PPVT			Math		Language		
							Mother Tongue (MT)		English
Round (Age)	Round 3 (Age 8)	Round 4 (Age 12)	Round 5 (Age 15)	Round 3 (Age 8)	Round 4 (Age 12)	Round 5 (Age 15)	Round 3 (Age 8)	Round 4 (Age 12)	Round 5 (Age 15)
Model 1: OLS unadjusted									
Preschool coeff.	1.19***	1.14***	1.14***	1.17***	0.92***	0.98***	0.76***	0.88***	1.00***
(SE)	(0.16)	(0.17)	(0.14)	(0.17)	(0.16)	(0.17)	(0.16)	(0.14)	(0.12)
Model 2: OLS with all covariates									
Preschool coeff.	0.43***	0.34***	0.24***	0.38***	0.02	0.04	0.15*	0.12	0.24***
(SE)	(0.12)	(0.08)	(0.08)	(0.11)	(0.10)	(0.05)	(0.10)	(0.11)	(0.10)
R-squared	0.46	0.46	0.46	0.47	0.32	0.33	0.31	0.33	0.35
Observations	1447	1417	1447	1417	1354	1447	1444	1320	1447
Model 3: Kernel matching estimates (primary approach)									
ATT	0.36***	0.21***	0.19***	0.51***	0.09	0.16*	0.29***	0.00	0.33***
(SE)	(0.12)	(0.10)	(0.07)	(0.11)	(0.20)	(0.11)	(0.12)	(0.11)	(0.15)
Average of treated	0.66	0.67	0.70	0.61	0.45	0.41	0.29	0.43	0.51
Average of matched control	0.30	0.46	0.51	0.10	0.36	0.24	-0.01	0.43	0.18
Observations	1447	1417	1447	1417	1354	1447	1444	1320	1447
Model 4: Kernel matching within sentinel site (robustness check)									
ATT	0.57***	0.25**	0.21***	0.50***	0.17	0.21**	0.21**	-0.08	0.08
(SE)	(0.15)	(0.14)	(0.14)	(0.16)	(0.17)	(0.15)	(0.16)	(0.18)	(0.17)
Observations	1032	1003	1032	1017	953	1032	1029	927	1032

Note: (1) Preschool coefficients (coeff.) and ATTs are based on the standardised score (z-score) of each test; (2) In Model 1 and 2: robust standard errors, clustered at community level, in parentheses; (3) In Model 3: standard errors using bootstrap (1,000 replications) in parentheses; (4) In Model 4 using kernel matching within sentinel site, six sentinel sites dropped due to the lack of either preschool attendees or non-attendees.
*** p<0.01, ** p<0.05, *p<0.1
Source: Young Lives Dataset Round 2 to Round 5, Young Lives

Table 4.7 shows the results of the same set of regression models when the analyses were restricted to the urban sample. The ATT estimates of the kernel-based matching (Model 3) showed similar patterns of the statistical significance with those from the pooled sample, whilst the estimates declined slightly across the subjects and ages. Consistent with the findings from the pooled sample, significant associations between preschool attendance and students' performance were observed in PPVT (0.40 *SD*, $p < 0.01$) and English (0.24 *SD*, $p < 0.05$) by the age of 15, while no significant associations were found in math and MT performance from age 12.

Table 4.7. Relation between Preschool and Academic Achievement, ATT (Urban Sample)

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
Subject		PPVT			Math		Language		
							Mother Tongue (MT)		English
Round (Age)	Round 3 (age 8)	Round 4 (age 12)	Round 5 (age 15)	Round 3 (age 8)	Round 4 (age 12)	Round 5 (age 15)	Round 3 (age 8)	Round 4 (age 12)	Round 5 (age 15)
Model 1: OLS unadjusted									
Preschool coeff.	0.92***	0.90***	0.88***	0.72***	0.49**	0.70***	0.43***	0.55**	0.83***
(SE)	(0.13)	(0.18)	(0.17)	(0.21)	(0.16)	(0.16)	(0.21)	(0.18)	(0.20)
Model 2: OLS with all covariates									
Preschool coeff.	0.24*	0.30**	0.30***	0.27***	0.03	0.05	0.06	0.01	0.23**
(SE)	(0.11)	(0.10)	(0.06)	(0.07)	(0.16)	(0.07)	(0.13)	(0.08)	(0.09)
R-squared	0.39	0.35	0.34	0.35	0.20	0.30	0.22	0.24	0.34
Observations	652	632	652	646	622	652	650	610	652
Model 3: Kernel matching estimates (primary approach)									
ATT	0.30***	0.17*	0.40***	0.42***	0.16	0.17*	0.19*	-0.16	0.24**
(SE)	(0.14)	(0.14)	(0.11)	(0.14)	(0.20)	(0.11)	(0.16)	(0.12)	(0.13)
Average of treated	0.21	0.20	0.25	0.11	0.06	0.01	-0.01	0.06	0.18
Average of matched control	-0.09	0.03	-0.15	-0.31	-0.10	-0.19	-0.20	0.20	-0.06
Observations	652	632	652	646	622	652	650	610	652
Model 4: Kernel matching within sentinel site (robustness check)									
ATT	0.32***	0.38***	0.23**	0.31***	0.34**	0.23**	0.22**	-0.27**	0.27**
(SE)	(0.20)	(0.22)	(0.19)	(0.21)	(0.22)	(0.18)	(0.22)	(0.21)	(0.22)
Observations	652	632	652	646	622	652	650	610	652

Note: (1) Preschool coefficients (coeff.) and ATTs are based on the standardised score (z-score) of each test; (2) In Model 1 and 2: robust standard errors, clustered at community level, in parentheses; (3) In Model 3: standard errors using bootstrap (1,000 replications) in parentheses; (4) In Model 4 using kernel matching within sentinel site, six sentinel sites dropped due to the lack of either preschool attendees or non-attendees.
*** p<0.01, ** p<0.05, *p<0.1
Source: Young Lives Dataset Round 2 to Round 5, Young Lives

Supplementary analysis for students' outcomes on language test. The fadeout of preschool influence on math achievement by age 12 in Ethiopia was suggested by Vandemooretele (2018), and the current analysis confirmed this pattern by age 15. For the language assessments, however, an irregular pattern was observed between Rounds 4 and 5, as the predictive role of preschool faded out by age 12 but re-emerged at age 15. Although this pattern could simply be a function of subject difference—that is, mother tongue *versus* English—my supplementary analysis suggested that it could be in part attributable to measurement issues.

To elaborate, the test difficulty of MT assessments, which consist of reading and listening comprehension tasks, seems to be higher than Ethiopian students' average academic ability. Besides the assessments used in the main analysis, Young Lives administered some additional language tests in each round: oral reading fluency and word recognition tests in Round 3, and simple reading tests in three languages (Amharic, MT, and English) in Round 4. Noticeably, in Round 3, although about one-third of students (34.6%, except missing responses) were not able to read any words on the oral reading fluency test, only 3 percent of students reported a zero score on the reading comprehension test.

Using these additional language tests, the ATT estimates of the kernel-based matching showed that the association between preschool attendance and oral reading fluency in Round 3 (age 8) was statistically significant (0.35 SD, $p < 0.01$) and that preschool attendees read 10.2 more words correctly per minute than their matched peers who did not attend preschool.[146] This suggests that, for 8-year-olds, preschool attendance is more strongly predictive of students' oral reading fluency skills than their reading and listening comprehension skills. In another reading test in Round 4 (age 12), which assessed students' reading skills in five categories—cannot read, read only letter, read only word, read single sentence, and read multiple sentences—preschool attendees were 25 percentage points more likely to read sentences in their mother tongue than non-attendees ($p < 0.01$) and 10 percentage points more likely to read words or sentences in English than non-attendees ($p < 0.05$).[147] The results of this supplementary analysis can explain in part why the predictive role of preschool seems to be weaker in MT assessments at age 8 and 12 but becomes stronger again in English assessments at age 15. With relatively easier test instruments, attending a preschool remained a significant contributor to students' MT performance across different ages.

Preschool and educational attainment. As shown in Table 4.8, I continued to investigate the relation between preschool and educational attainment as measured by the highest grade achieved and on-time grade progression at ages 12 and 15. According to the ATT estimates of the kernel-based matching (Model 3) for the pooled sample, preschool attendees completed 0.74 higher grades ($p < .01$) than non-attendees. Students who attended preschool also had a 27.1 percentage point higher chance of adequately progressing by grade by age 15.[148] Among the urban sample (last four columns), preschool attendees completed 0.33 higher grades ($p <$

---

[146] The results on the oral reading fluency present in the Appendix Table E.1 and E.2.

[147] For 12-year-olds, the reading level categories for MT were re-grouped into (i) cannot read at all; (ii) read only letter or words; and (ii) read single or multiple sentences; for English (15-year-olds) (i) cannot read at all; (ii) read only letter; and (iii) read words or sentences.

[148] On-time progression was calculated based on Young Lives students' ages (age 14 or 15). In Round 5 (n=1,447), 525 students were age 14 and 922 students were age 15, thus those who reported the completion of Grade 7 or Grade 8 were counted for 'on-time grade progression'. Besides, secondary school transition needs to be considered only for students age 15 (n=922 of the sample) who reported the completion of Grade 8—for example, among 15-year-olds, those who attended preschool had a 17.4% higher chance of progressing to secondary school at the proper age.

.01) than non-attendees and had a 15 percentage point higher chance of adequately progressing by grade by age 15 (p < .01). The ATT estimates applying the kernel matching within the sentinel (Model 4) showed patterns consistent with Model 3, except students' outcomes in Round 4 among the urban sample.

Table 4.8. Relation between Preschool and Educational Attainment, ATT (Pooled/Urban)

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
	Pooled Sample				***Urban Sample***			
Education Attainment	Highest Grade Achieved		On-time Progression		Highest Grade Achieved		On-time Progression	
Round (Age)	Round 4 (age 12)	Round 5 (age 15)	Round 4 (age 12)	Round 5 (age 15)	Round 4 (age 12)	Round 5 (age 15)	Round 4 (age 12)	Round 5 (age 15)
Model 1: OLS unadjusted			*(Odds ratio)*				*(Odds ratio)*	
Preschool coeff. /Odds ratio	0.99***	1.20***	*2.46****	*4.08****	0.61***	0.68***	*1.49***	*2.89****
(SE)	(0.30)	(0.31)	*(0.62)*	*(1.32)*	(0.42)	(0.48)	*(0.70)*	*(1.48)*
Model 2: OLS with all covariates			*(Odds ratio)*				*(Odds ratio)*	
Preschool coeff. /Odds ratio	0.62***	0.56***	*2.04***	*3.39****	0.40***	0.37***	*1.82***	*2.64****
(SE)	(0.15)	(0.18)	*(0.60)*	*(0.66)*	(0.18)	(0.24)	*(0.72)*	*(0.68)*
R-squared	0.43	0.38	-	-	0.38	0.37	-	-
Observations	1402	1,447	1,447	1447	649	652	652	652
Model 3: Kernel matching estimates (primary approach)								
	Grade		%		Grade		%	
ATT	0.75***	0.74***	0.30***	0.27***	0.35***	0.33**	0.20***	0.15***
(SE)	(0.21)	(0.23)	(0.07)	(0.07)	(0.19)	(0.26)	(0.07)	(0.08)
Average of treated	5.31	7.06	0.67	0.73	5.42	7.14	0.69	0.76
Average of matched control	4.56	6.32	0.37	0.46	5.07	6.81	0.49	0.61
Observations	1402	1447	1447	1447	649	652	652	652
Model 4: Kernel matching within sentinel site (robustness check)								
	Grade		%		Grade		%	
ATT	0.52***	0.57***	0.17***	0.23***	0.09	0.40**	0.08	0.17***
(SE)	(0.26)	(0.29)	(0.09)	(0.08)	(0.29)	(0.34)	(0.11)	(0.11)
Observations	1014	1032	1032	1032	649	652	652	652

Note: (1) Italics are 'odds ratio'; (2) In Model 1 and 2: robust standard errors, clustered at community level, in parentheses; (3) In Model 3: standard errors using bootstrap (1,000 replications) in parentheses; (4) In Model 4 using kernel matching within sentinel site, six sentinel sites dropped due to the lack of either preschool attendees or non-attendees.
*** p<0.01, ** p<0.05, *p<0.1
Source: Young Lives Dataset Round 2 to Round 5, Young Lives

4.7.2 Research Question 2: Variation in Preschool Influence by Child and Family Characteristics

Table 4.9 presents the differential influence of preschool by sub-groups, as defined by child gender, household wealth, father's education level, and child's prior achievement level. To avoid cluttering the tables, Table 4.9 shows the ATT estimates of the kernel-based matching only (correspond to Model 3 in Tables 4.6 to 4.8) and outcome variables measured at age 15 (Round 5), including PPVT, English, and highest grade achieved. This table illustrates the mean scores/grades by treated and matched control group, ATT estimates, and standard errors for the matched sample. The upper and lower parts of the table show the pooled and urban sample, respectively.

Preschool and gender. The influence of preschool varied by gender. It showed significant benefits for both boys and girls on PPVT achievement, but greater benefits among boys on language achievement and highest grade achieved by age 15. Boys who attended preschool were scored an average of 0.20 *SD* higher in PPVT ($p < 0.05$), 0.53 *SD* higher in English ($p < 0.01$), and they completed 1.03 additional school years ($p < 0.01$) by age 15, all of which were statistically significant. For girls the benefits of preschool were significant in PPVT by an average of 0.18 *SD* ($p < 0.05$) and in grade completion by 0.40 additional year ($p < 0.05$). Similar patterns of gender difference were observed in the urban sample. I also estimated the model that included the interaction term between preschool and gender, which turned out to be not statistically significant, meaning that differences between preschool boys and preschool girls were similar to those between non-preschool boys and non-preschool girls.

Preschool and household wealth. The positive influence of preschool appears to be greater for students from more affluent households than for those from poorer households, divided by the median of household wealth index. For students from the upper wealth group, test scores increased by 0.21 *SD* in PPVT and 0.36 *SD* in English ($p < 0.01$ for both), and they completed 0.78 additional school years on average ($p < 0.01$). In contrast, for the lower wealth group there was no significant benefit from preschool across three outcome measures.[149] Among the urban sample, rich children consistently benefited more from preschool on their achievement in PPVT and English, whereas the benefit on educational attainment was more pronounced for poor children; that is, they completed 0.54 additional years ($p < 0.01$) than their peers who did not attend preschool. The model with the interaction term between preschool and household wealth was found to be not statistically significant across the three outcome variables.

Preschool and father's education level. The influence of preschool varies by father's education level, which was categorised by no education, some primary education or completion, and secondary education and above. The gains from preschool tended to be greater

[149] I also estimated the differential benefits of preschool by household wealth tercile (3 groups) and limited my sample to only two upper wealth groups (given that preschool attendance is very low in the lowest wealth group). The estimates from this supplementary analysis consistently indicated that, within the two upper wealth groups, children from richer households obtained greater benefits than those from poorer households.

for students with fathers who enrolled in or completed primary education. For students from this middle group, PPVT and English test scores increased by 0.25 *SD* and 0.41 *SD* ($p < 0.01$ for both), and they completed 0.85 additional grades ($p < 0.01$) by age 15. For students from the lowest group (no education), the preschool benefit was more pronounced only for English achievement (0.62 *SD*, $p < 0.05$). For students from the highest group (secondary education and above), although there were no significant preschool benefits for academic achievement, they completed 0.77 additional school years ($p < 0.01$) by age 15 than their peers with no preschool experience. Similar patterns—that is, greater gains for the middle group—were observed for the urban sample, while children from all three groups obtained significant gains from preschool on their achievement in PPVT (0.35 *SD* to 0*.51 SD*, $p < 0.01$). The model that considered the interaction term between preschool and father's education level turned out to be not statistically significant in all outcome variables.

Preschool and prior achievement level. When grouped by students' prior achievement, the positive influence of preschool appeared to be greater for students from the middle achievement group, followed by those from the low achievement group. For students from the middle achievement group, those who attended preschool were scored an average of 0.56 *SD* higher in PPVT, 0.76 *SD* higher in English, and they completed 1.64 additional school years by age 15, all of which were statistically significant at the 0.01 level. Similarly, students from the low achievement group scored 0.46 *SD* ($p < 0.01$) higher in English and completed 0.98 additional grades ($p < 0.01$) by the age of 15, although no significant benefits were found on PPVT achievement. In contrast, there was no significant benefit for students from the high achievement group across all three outcomes. It may be worth noting that children from the middle achievement level showed some extremes in their later achievement, depending on preschool attendance. Preschool attendees in this middle group had similar or even higher achievement than those from the high achievement group, whereas preschool non-attendees in this group had significantly lower achievement than those from the low achievement group. Among the urban sample, the two lower groups consistently gained the most from preschool on academic achievement, the one exception being on higher educational attainment by the upper achievement group (0.58 additional grades, $p < 0.01$). The model with the interaction term between preschool and students' prior achievement was found to be not statistically significant across the three outcome variables.

Table 4.9. Sub-Group Analysis by Child and Family Characteristics, ATT (Pooled/Urban)

Subject, Round (age)	PPVT, Round 5 (age 15)				Language (English), Round 5 (age 15)				Highest Grade Achieved, Round 5 (age 15)			
	Treated	Matched control	*ATT*	*(SE)*	Treated	Matched control	*Coeff.*	*(SE)*	Treated	Matched control	*Coeff.*	*(SE)*
					Pooled Sample							
Preschool attendance	0.70	0.51	0.19***	(0.08)	0.51	0.18	0.33***	(0.18)	7.06	6.32	0.74***	(0.28)
A. Gender												
Boys	0.70	0.50	0.20**	(0.11)	0.43	-0.10	0.53***	(0.20)	6.90	5.87	1.03***	(0.36)
Girls	0.71	0.53	0.18**	(0.10)	0.60	0.50	0.10	(0.27)	7.27	6.87	0.40**	(0.30)
B. Household Wealth												
Lowest 50 percent	0.30	0.18	0.12	(0.24)	0.15	-0.04	0.19	(0.23)	6.35	5.67	0.69	(0.31)
Top 50 percent	0.77	0.56	0.21***	(0.08)	0.56	0.20	0.36***	(0.20)	7.18	6.40	0.78***	(0.31)
C. Father's Education Level												
No education	0.64	0.45	0.19	(0.16)	0.38	-0.24	0.62**	(0.38)	6.63	6.33	0.30	(0.41)
Some primary/ completion	0.64	0.39	0.25***	(0.10)	0.52	0.11	0.41***	(0.14)	6.94	6.09	0.85***	(0.53)
Secondary and above	0.80	0.69	0.11	(0.10)	0.54	0.43	0.11	(0.35)	7.38	6.61	0.77***	(0.23)
DE. Prior Achievement level												
Low	0.53	0.36	0.17	(0.13)	0.56	0.10	0.46**	(0.20)	7.33	6.35	0.98***	(0.41)
Middle	0.79	0.23	0.56***	(0.10)	0.56	-0.20	0.76***	(0.20)	6.89	5.25	1.64***	(0.55)
High	0.71	0.68	0.03	(0.09)	0.45	0.37	0.08	(0.27)	7.07	6.82	0.25	(0.25)
					Urban Sample							
Preschool attendance	0.25	-0.15	0.40***	(0.12)	0.18	-0.05	0.23**	(0.14)	7.14	6.81	0.33**	(0.32)
A. Gender												
Boys	0.25	-0.13	0.38***	(0.21)	0.09	-0.26	0.35***	(0.20)	6.94	6.16	0.78***	(0.42)
Girls	0.26	-0.15	0.41***	(0.12)	0.31	0.15	0.16	(0.18)	7.42	7.43	-0.01	(0.33)
B. Household Wealth												
Lowest 50 percent	0.10	-0.26	0.36***	(0.15)	0.07	-0.06	0.13	(0.22)	6.92	6.38	0.54**	(0.54)
Top 50 percent	0.37	-0.05	0.42***	(0.18)	0.28	-0.05	0.33***	(0.17)	7.31	7.18	0.13	(0.32)
C. Father's Education Level												
No education	0.13	-0.38	0.51***	(0.21)	0.04	0.00	0.04	(0.40)	6.79	7.02	-0.23	(0.41)
Some primary/ completion	0.11	-0.24	0.35***	(0.20)	0.18	-0.24	0.42***	(0.18)	6.96	6.38	0.58**	(0.59)
Secondary and above	0.45	0.10	0.35***	(0.14)	0.24	0.19	0.05	(0.19)	7.46	7.31	0.15	(0.26)
D. Prior Achievement Level												
Low	0.08	-0.32	0.40***	(0.18)	0.20	-0.10	0.30**	(0.16)	7.04	6.70	0.34	(0.31)
Middle	0.27	-0.30	0.57***	(0.15)	0.22	-0.06	0.28*	(0.26)	7.09	6.91	0.18	(0.67)
High	0.46	0.38	0.08	(0.19)	0.10	0.02	0.08	(0.21)	7.33	6.74	0.58**	(0.29)

*** p<0.01, ** p<0.05, *p<0.1 *Source:* Young Lives Dataset Round 2 to Round 5, Young Lives

4.7.3 Research Question 3(1): Variation in Preschool Influence by Preschool Characteristics

Table 4.10 shows results stratified by various characteristics of preschool: (1) preschool starting age (4, 5, and 6)[150]; (2) preschool type (private, government-funded, and community-based); (3) subjective preschool quality reported by parents (okay or bad, good, and excellent); and (4) daily hours of participation (half or full day). Considering the limited measures for preschool quality, my examinations here are considered exploratory. The upper and lower parts of the table represent the pooled and urban samples, respectively.

Preschool starting age. At the top of Table 4.10, I separately show the ATT estimates of kernel matching for students starting preschool at age 4, 5, or 6 (or later) *versus* students who never attended preschool. As for academic achievement, early participation in preschool at ages 4 and 5 was associated with significantly improved scores in PPVT (0.23 *SD* for age 4 and 0.21 *SD* for age 5, $p < 0.01$) and English achievement (0.35 SD for both ages, $p < 0.1$), but there was no significant association with later participation in preschool at age 6. Regarding educational attainment, early and late participation in preschool were both associated with the highest grade achieved at age 15. For all three sub-groups, the preschool benefits were nearly equivalent to three-quarters of an additional school year by age 15. Similarly, within the urban sample, entrance to preschool at age 4 led to greater gains in academic achievement in PPVT (0.57 SD, $p < 0.01$) and English (0.37 SD, $p < 0.01$). No significant differences were observed for the outcome of the highest grade achieved, based on the age of first preschool participation.

Preschool type. I compared students attending different types of preschool to students who did not attend any type of preschool. Since private kindergarten was the dominant service provider in Ethiopia before 2010, significant and positive associations were found only between private-run preschools and students' outcomes in PPVT (0.29 *SD*, $p < 0.01$), language achievement (0.41 *SD*, $p < 0.05$), and the highest grade achieved (0.79 additional school year, $p < 0.01$). One exception was the significantly strong association between government-funded preschool and students' highest grade achieved. The benefit of attending a government-funded preschool was equivalent to more than one single year of schooling (1.23 additional school year, $p < 0.01$)

[150] Duration of preschool attendance was considered; however, it was hard to measure, given that 406 of 476 (or 85.3%) children were still enrolled in preschool when the Round 2 survey was conducted.

and greater than attending a private-run preschool (0.86 additional school year, $p < 0.01$). This pattern continued in the urban sample, as represented by the greater benefit of private preschool on academic achievement and the larger gains from government preschool on educational attainment.

Preschool quality. On the measure of preschool quality, significant and positive associations were observed between high-quality preschool and students' PPVT test scores (0.32 *SD*, $p < 0.01$), but it was not significantly associated with students' English test scores. Regardless of the quality of preschool, all preschool attendees completed higher grades than those who never attended preschool. Similarly, among the urban sample, strong associations appeared only between high-quality preschool and students' PPVT test scores. As explained previously, the measure of preschool quality could be subjective, as it relied on the self-reports of parents who sent their children to preschool. They most likely believed it was worthwhile to pay for preschool, as it provided good-quality services that enhanced child development.

Daily hours of preschool participation. Regarding how long children attended preschool daily, significant associations between preschool attendance and academic achievement (0.31 *SD* for PPVT, $p < 0.01$) were observed only for full-time preschool, which offered more than seven hours of services per day. Both groups attending half-day or full-day preschool achieved higher grades than those who did not attend preschool, by 0.63 and 0.80 additional schooling, by the age of 15. Within the urban sample, significant and positive associations were found only between full-time preschool and students' PPVT outcomes. Although there were some variations, the results focusing on PPVT outcomes confirmed the hypothesis that preschool, which offers high-quality and more intensive exposure, may relate to more favourable outcomes for children.

Table 4.10. Sub-Group Analysis by Preschool Quality, ATT (Pooled/Urban)

Subject	PPVT		Language (English)		Highest Grade Achieved	
Round (age)	Round 5 (age 15)		Round 5 (age 15)		Round 5 (age 15)	
	Pooled Sample					
	ATT	(SE)	ATT	(SE)	ATT	(SE)
A. Preschool starting age						
Age 4	0.23**	(0.11)	0.35*	(0.21)	0.75**	(0.31)
Age 5	0.21***	(0.08)	0.35*	(0.19)	0.69**	(0.29)
Age 6 or later	0.11	(0.11)	0.26	(0.21)	0.83**	(0.33)
B. Preschool type						
Private	0.29***	(0.08)	0.41**	(0.19)	0.79***	(0.28)
Government-funded	-0.20	(0.22)	0.10	(0.30)	1.23***	(0.37)
Community-based	-0.03	(0.11)	0.12	(0.21)	0.39	(0.32)
C. Subjective preschool quality						
Okay or Bad	0.09	(0.11)	0.35*	(0.20)	0.86***	(0.30)
Good	0.20**	(0.08)	0.33*	(0.19)	0.63**	(0.29)
Excellent	0.32***	(0.09)	0.28	(0.21)	0.89***	(0.31)
D. Daily hours of preschool participation						
Half day	-0.04	(0.10)	0.27	(0.20)	0.63**	(0.30)
Full day (> 7hrs)	0.31***	(0.08)	0.36*	(0.19)	0.80***	(0.28)
	Urban Sample					
	ATT	(SE)	ATT	(SE)	ATT	(SE)
A. Preschool starting age						
Age 4	0.57***	(0.15)	0.37**	(0.17)	0.45	(0.35)
Age 5	0.34***	(0.13)	0.19	(0.15)	0.26	(0.33)
Age 6	0.37**	(0.18)	0.21	(0.18)	0.38	(0.37)
B. Preschool type						
Private	0.48***	(0.12)	0.29**	(0.14)	0.35	(0.33)
Government-funded	-0.09	(0.32)	-0.07	(0.27)	1.08***	(0.36)
Community-based	0.22	(0.19)	0.13	(0.20)	-0.24	(0.40)
C. Subjective preschool quality						
Okay or Bad	0.36**	(0.16)	0.31*	(0.16)	0.53	(0.34)
Good	0.35***	(0.13)	0.19	(0.15)	0.16	(0.33)
Excellent	0.60***	(0.13)	0.23	(0.18)	0.46	(0.37)
D. Daily hours of preschool participation						
Half day	0.07	(0.16)	0.25	(0.16)	0.34	(0.35)
Full day (> 7hrs)	0.54***	(0.12)	0.23	(0.14)	0.33	(0.33)

*** p<0.01, ** p<0.05, *p<0.1
Source: Young Lives Dataset Round 2 to Round 5, Young Lives

4.7.4 Research Question 3(2): Mediating Role of Subsequent School Experience

This section further explores how subsequent school experience mediates the relationship between preschool and students' academic achievement. Aligned with the previous chapter, I applied mediation analysis using the structural equation modeling (SEM) framework to address this research question (see Section 3.6.2. for empirical strategy).

1) Data: Primary school characteristics after pre-primary education

The mediation analysis in the present study used the dataset from the Young Lives Ethiopia school survey, which was administered in 2012-2013 separately from the household-based Young Lives survey. The Young Lives school survey 2012-2013, conducted at the beginning and end of the academic year, was designed to allow researchers to investigate what shapes children's learning and progression over the course of a school year. The school survey provides a wealth of information about pupils (e.g., academic achievement and school engagement), teachers (e.g., teacher efficacy and pedagogical content knowledge), and schools (e.g., school resources and organisation) (see Aurino, James, & Rolleston, 2014, for further details).[151] One strength of the Young Lives school survey is that some information was collected via classroom observation conducted by field workers, rather than relying on principals' self-reported responses. For instance, the school asset index was constructed based on the physical resources in schools that were collected by direct observation, such as having electricity, a functional library, internet access, working computers, a sports or play area, working latrines/toilets, water facilities for various uses, etc.

The original sample of the Young Lives school survey 2012-2013 was nearly 12,000 primary school students studying in Grades 4 and 5. The school survey 2012-2013 was conducted at 30 sites, including the 20 main Young Lives sites (in the regions of Addis Ababa, Amhara, Oromia, SNNP, and Tigray) and 10 additional sites from two historically disadvantaged regions of Ethiopia, Somali and Afar.[152] Crucially for the purpose of the present study, I used only the sample overlapping between the main Young Lives survey and the school survey 2012-2013 in order to capture students' educational trajectories from early childhood to adolescence. Therefore, the study sample was limited to the 549 students who straddled the longitudinal household survey and the cross-sectional school survey. This final sample is approximately one-quarter of the Young Lives Younger Cohort sample and about 5 percent of the school survey 2012-2013 sample.

[151] The third round of the Young Lives school survey was administered in 2016-2017, but it was not included in the present study due to the limited size of overlap with the main Young Lives survey.
[152] The School Survey 2012-2013 constituted a site-level census, as it sampled all pupils (including both Young Lives children and non-Young Lives children) studying in all Grade 4 and Grade 5 classes in all schools located within the geographic boundaries of each survey sentinel site (Aurino, James, & Rolleston, 2014).

2) Descriptive statistics and SEM models

Descriptive statistics for school characteristics drawn from the Young Lives school survey 2012-2013 are presented in Table 4.11.[153] Most of the schools in the survey were public and had an average school asset index of 0.73. Only one-quarter of the schools had classrooms with fewer than 50 students. While 63 percent of the schools showed a 1:1 textbook-pupil ratio, about one-third used a newly revised government MT textbook. About one-third of schools reported that principals provided some incentives for well-performing teachers, such as financial incentives, promotion to a higher level of teaching, teacher training courses, or certificates, prizes, or ceremonies.

Table 4.11. Descriptive Statistics: Selected School Characteristics

School characteristics	**School Survey 2012-2013**	
	Mean	(SD)
(1) School asset index	0.73	(0.21)
(2) Average classroom size, less than 50	0.27	(0.44)
(3) 1:1 MT textbook-pupil ratio	0.63	(0.48)
(4) Use of new government language textbook	0.32	(0.47)
(5) Incentive for a well-performing teacher	0.33	(0.47)
(6) Low school attendance	0.22	(0.42)
(7) School is public school.	0.96	(0.19)
Observation (Students)	549	

Note: (1) School asset index is calculated by the sum of '1 (yes)' from binary indicators divided by the number of items (e.g., radio, computer, library, internet, working toilets, etc). (2) School asset index, average classroom size, textbook-pupil ratio, and low school attendance were collected via observation by field workers; the rest is from principal's response.
Source: Young Lives School Survey 2012/13, Young Lives

Table 4.12 shows the results of pairwise correlation analysis among the hypothesised mediators of school characteristics to investigate how these variables are correlated with outcome (dependent) variables. Except for two measures (class size, availability of a new government textbook), five measures of school characteristics were moderately correlated with PPVT test scores at ages 12 and 15, which ranged from 0.16 to 0.30. Given that the majority of schools were public, I decided to include four measures—school asset index, MT textbook-pupil ratio, students' attendance level, and incentives for a well-performing teacher—in order to construct an indicator of subsequent schooling environments. The estimates from the *t-test* also indicated a statistically significant difference between preschool attendees and non-attendees in the four variables ($p < 0.01$ for school assets, students' attendance; $p < 0.05$ for textbook-pupil ratio,

[153] Given the exhaustive list of school (class and teacher) characteristics in the Young Lives School Survey 2012-2013, only selected variables are presented in Table 4.11.

teacher incentives). While the former two measures captured the schools' structural quality, such as infrastructure and learning materials, the last measure on teacher incentive was relevant to process quality that could facilitate the interaction between students, teachers, and principals.

In prior studies on the effectiveness of education interventions in LMICs, evidence varied by specific school-level characteristics, target populations, or other contextual factors (Conn, 2014). For instance, according to the review articles summarizing the estimated effects of school materials interventions (e.g., textbooks, writing materials, and flipcharts) on students' learning outcomes, school materials seemed to have a significant and positive effect on students' math scores but not on language (Krishnaratne et al., 2013), or to be compounded by effects of co-occurring teacher training and class size reduction programmes (McEwan, 2014). Evidence on textbook distribution was more mixed, as it could interact with numerous contextual factors, including students' basic literacy levels, varying levels of difficulty of the textbooks, or differences in how teachers integrated these books into the curriculum (Conn, 2014). An individual randomised evaluation of textbook provision showed it only had an impact for high-performing students in Kenya (Glewwe et al., 2009) and for students living in rural areas across five francophone countries (Frölich & Michaelowa, 2011). Similarly, teacher incentives produced relatively small or non-significant effects on student outcomes across ten studies conducted in LMICs (Snilstveit et al., 2016), with one exception conducted in Kenya (Duflo, Dupas, & Kremer, 2015). In this study, teacher incentives included offering short-term teacher contracts to local teachers, which led to significantly improved student learning outcomes only when combined with class size reduction (Duflo et al., 2015). By comparison, the teacher incentive indicator used in the present study entailed both formal and informal rewards for teachers (e.g., financial incentive, job promotion, or a certificate or prize) that were likely offered by school leadership. Although there is no clear guidance from prior studies to generate a hypothesis, the four measures selected for the current analysis have been frequently used in the existing literature.

Table 4.12. Pairwise Correlation for Mediating and Dependent Variables

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
(1) School asset	1.00								
(2) Class size	-0.01	1.00							
(3) Text ratio	0.27***	-0.12***	1.00						
(4) New Text	-0.16***	-0.01	-0.01	1.00					
(5) Incentive	0.12***	0.04	0.10**	-0.02	1.00				
(6) Low Att.	-0.24***	0.06	-0.16***	0.21***	-0.23***	1.00			
(7) School Type	-0.01	-0.31***	0.01	-0.11**	0.02	0.08	1.00		
(8) PPVT-R4	0.21***	-0.04	0.30***	0.02	0.16***	-0.30***	-0.17***	1.00	
(9) PPVT- R5	0.18***	-0.02	0.16***	0.01	0.23***	-0.25***	-0.17***	0.71***	1.00

*** p<0.01, ** p<0.05, *p<0.1
Source: Young Lives Dataset Round 2 to Round 5, Young Lives

After identifying the potential mediators, I estimated two sets of SEM models to determine the best fitting model for the present data. Model 1 first included preschool attendance as a direct predictor of four observed measures of school environment (i.e., school asset index, MT textbook-pupil ratio, students' attendance level, and incentives for a well-performing teacher), and students' PPVT scores. Next, these four measures of school environments were included as direct predictors of students' PPVT scores. To account for common sources of measurement error, error terms of the four variables representing school environments were allowed to be correlated. In Model 2—represented visually in Figures 4.3 and 4.4—I introduced the *latent* variable of school environments based on the selected observed variables. Model 2 included preschool attendance as a direct predictor of the latent measure of school environments and students' PPVT scores, then this latent measure was included as a direct predictor of students' PPVT scores. On the latent measure of school environments, I conducted a one-factor confirmatory factor analysis and confirmed that this shows an adequate model fit ($\chi^2(2) = 5.89$, $p < 0.001$; CFI = 0.93; RMSEA = 0.06).

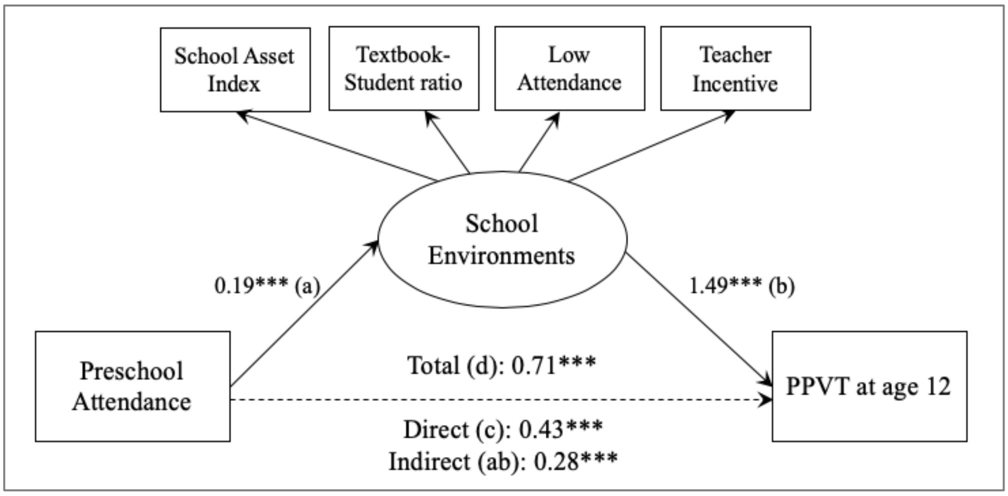
Across all models, covariates were included to account for potential sources of selection bias. All covariates, including child gender, father's education level, household size, whether a child was the first born, language between home and school, and wealth quintile, had paths to preschool attendance, mediators, and students' outcomes. Analyses were conducted in Stata version 14.1. For all models, adequate model fit was indicated by a root mean squared error of approximation (RMSEA) of ≦ 0.06 (Hu & Bentler, 1999) and a comparative fit index (CFI) of ≧ 0.90 (Bentler, 1990). The traditional goodness-of-fit statistic (i.e., a nonsignificant chi-square) test was relaxed because the 'chi-square value can be overly influenced by sample size, correlations, variance unrelated to the model, and multivariate non-normality' (Kline, 2011, p. 201).

3) Results of SEM

Using the mediation analysis framework, I examined the extent to which school-level characteristics mediated the relations between preschool attendance and students' receptive vocabulary skills during adolescence. Overall, SEM Model 2, in which a latent variable was included, showed adequate model fit statistics: (with PPVT-Round 5; $\chi^2(26) = 64.55$; CFI = 0.89; RMSEA = 0.05) (see Table 4.13). Compared to Model 2, Model 1 (which included four observed variables of school environment) showed a significantly poorer fit (with PPVT-Round 5; $\chi^2(3) = 30.28$; CFI = 0.92; RMSEA = 0.14; SRMR = 0.03) (see Appendix Table E.4). Thus, I focused on presenting and interpreting the results of Model 2 as the final model.

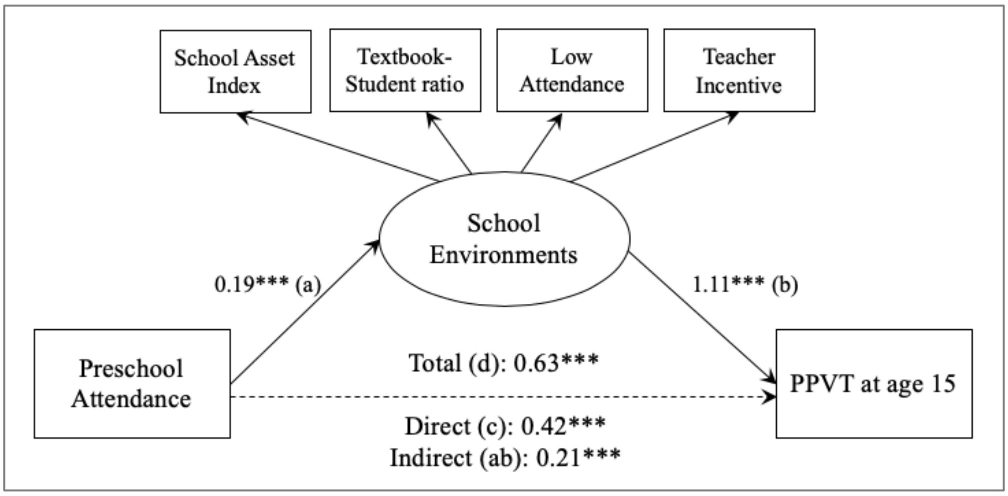
From the SEM results of Model 2, standardised coefficients of the total, direct, and indirect effects on PPVT Rounds 4 (age 12) and 5 (age 15) are presented in Figures 4.3 and 4.4. To ease the interpretation, I use the term 'effect' in the mediation analysis, where effect indicates 'association'. At the age of 12, preschool attendance positively influenced the mediator of schools' characteristics ($\beta = 0.19$, SE = 0.06, $p < 0.001$). Preschool attendance was also a positive predictor (direct effect) of PPVT scores in the model ($\beta = 0.43$, SE = 0.12, $p < 0.001$), after controlling for child gender, father's education level, household size, whether a child was the first born, language between home and school, and wealth quintile. In addition to the direct effects, I assessed the indirect effects among the study variables. The indicator of schools' environments was found to be a statistically significant mediator ($\beta = 0.28$, SE = 0.10, $p < 0.001$). Similar relationships were observed for students' PPVT outcomes at age 15. Preschool attendance positively influenced the mediator of schools' characteristics ($\beta = 0.19$, SE = 0.06, $p < 0.001$). Preschool attendance was also a positive predictor (direct effect) of PPVT scores in the model ($\beta = 0.42$, SE = 0.11, $p < 0.001$), after accounting for a set of covariates. In terms of the indirect effects among the study variables, the indicator of schools' environments was found to be a statistically significant mediator ($\beta = 0.21$, SE = 0.07, $p < 0.001$).

Figure 4.3. SEM Model on Preschool Attendance and PPVT at Age 12



Note: (Total Effects (d, 0.71) = Indirect Effects (a*b, 0.1898911*1.490002) + Direct Effects (c, 0.43).
*** p<0.01, ** p<0.05, *p<0.1
Source: Young Lives Dataset Round 2 to Round 5, Young Lives School Survey Dataset 2012-2013, Young Lives

Figure 4.4. SEM Model on Preschool Attendance and PPVT at Age 15



Note: (Total Effects (d, 0.63) = Indirect Effects (a*b, 0.1889541 * 1.10886) + Direct Effects (c, 0.42).
*** p<0.01, ** p<0.05, *p<0.1
Source: Young Lives Dataset Round 2 to Round 5, Young Lives School Survey Dataset 2012-2013, Young Lives

Table 4.13 summarizes the results of the SEM model. Given that no single index appears to be a viable mediation effect size measure (Wen & Fan, 2015), I followed the recommendations of Sobel (1982) to use (1) the proportion of total effect that is mediated (i.e., total effect explained by the indirect effect); and (2) the ratio of the indirect effect to direct effect (R_m statistic) as a proxy measure for the magnitude of mediation effect. As shown in Table 4.13, the indirect path from preschool attendance to students' PPVT score at age 12 via schools' characteristics explained 39.4 percent of the total association between preschool attendance and ORF score

(β = 0.71), and the ratio of the indirect effect to direct effect is 0.65. In addition, the indirect path mediated by schools' environments explained 33.3 percent of the total association between preschool attendance and students' PPVT score at age 15 (β = 0.63), with the ratio of the indirect effect to direct effect at 0.50. The contribution of the indirect effects is about half of the direct effects in PPVT at age 12 (R_m = 0.65) and at age 15 (R_m = 0.50), attenuated as students make the transition from primary to secondary. In this model, the mediator of schools' environment explained about one-third of the total effects for the association between preschool attendance and students' outcomes at ages 12 and 15.

Collectively, the results of the SEM suggest that the relations between preschool attendance and students' achievement at ages 12 and 15 were partially accounted for by subsequent schooling experience. There are several limitations of the current analysis, including a restricted sample, partially captured school environments at one point in time, and a single-level SEM framework. Despite these limitations, the findings of this study enhance our understanding of the mechanisms underlying students' skill formation from early childhood to adolescence and the environmental factors that could potentially affect the sustained benefits of preschool up to age 15.

Table 4.13. Results of SEM on the Link between Preschool Attendance and PPVT

	(1) PPVT - Round 4	(2) PPVT- Round 5
Age	(age 12)	(age 15)
Path coefficients **(S.E)**		
Preschool attendance → School Environments → PPVT outcome		
Total effects	0.71*** (0.11)	0.63*** (0.10)
Direct effects	0.43*** (0.12)	0.42*** (0.11)
Indirect effects	0.28*** (0.10)	0.21*** (0.07)
% of total effect mediated	39.4	33.3
% of total effect unmediated	60.6	66.7
Ratio of indirect effect to direct effect (R*m*)	0.65	0.50
Model Fit		
Chi-Square (*DF*)	70.88(26)	64.55(26)
CFI	0.87	0.89
RMSEA	0.05	0.05
SRMR	0.03	0.03
R-Square	0.31	0.34

Note: (1) The resulting structural coefficients (standardised regression coefficients) describe the direct and indirect effects. (2) Standard errors are in parentheses; (3) DF: Degree of Freedom; (4) CFI: Comparative Fit Index; (5) RMSEA: Root Mean Square Error of Approximation; (6) SRMR: Standardised Root Mean Square Residual.
*** p<0.01, ** p<0.05, *p<0.1
Source: Young Lives Dataset Round 2 to Round 5, Young Lives School Survey Dataset 2012-2013, Young Lives

4.8 Discussion

By drawing on data which tracked children in Ethiopia over 15 years, the present study has addressed gaps in the knowledge related to the lasting influence of preschool attendance on student outcomes in adolescence, particularly in a low-income context. Although the data captured the period when huge disparities in preschool access existed between advantaged and disadvantaged children, this study presents estimates using kernel-based matching (PSM) to adjust for several non-random sources of selection into preschool (e.g., household wealth, father's education, geographic location, child's nutrition and prior achievement). These estimates have been supported by alternative approaches, including OLS and kernel-based matching within specific community sites. While existing studies in Ethiopia have explored similar research questions using the Young Lives data (e.g., Woldehanna & Gebremedhin, 2012; Woldehanna & Araya, 2017; Vandemoortele, 2018), the current study extends that work with more comprehensive outcome measures at later ages, replicating the earlier findings with the Younger Cohort data, and considering additional hypotheses on key dimensions of student, family, and school characteristics.

Overall findings of the present study align with previous work suggesting a positive medium-term contribution of ECE to the cognitive skills, school enrolment, and educational progression of children in LMICs (Berlinski et al., 2008; Bietenbeck et al., 2017; Hazarika & Viren, 2013; Krafft, 2015). First, the results from the current study revealed that children who attended preschool between ages 4 and 6 not only outperformed their peers in academic achievement at age 8, but that these benefits carried forward over time; by age 15 they showed improved achievement in receptive vocabulary (0.19 *SD*) and English (0.33 *SD*). These results complement prior evidence from Sub-Saharan Africa showing the association between preschool attendance and 13- to 16-year-old's academic achievement in Kenya and Tanzania, as measured by the composite score of English, numeracy, and language tests (0.08 to 0.12 *SD*) (Bietenbeck et al., 2017). However, in the meantime, the present study found some variations by subject, in which a convergence of test scores between preschool attendees and non-attendees occurred at age 12 in math and mother tongue. Although this result differs from the evidence from India and Uganda that documents a positive significant association between preschool attendance and math achievement by age 11-12 (Hungi & Ngware, 2018; Singh & Mukherjee, 2018), the fadeout of preschool gains has been reported in Ethiopia, particularly on math achievement for 12-year-old students (Vandemoortele, 2018), which was confirmed by the current study up to age 15.

When we look closely at the mixed pattern by subjects, it is noteworthy that the sizable benefits of preschool appeared on students' achievement in PPVT up to age 15. Although the initial academic gains were attenuated over time—from 0.36 *SD* at age 8 to 0.19 *SD* at age 15—it remained significant at the 0.01 level. As supported by prior work in the U.S., some convergence in preschool effects is inevitable over the course of primary school, while it could be maintained by adolescence (Ansari, 2018; Bassok et al., 2018). By contrast, regarding the lack of observable benefits in math and mother tongue after age 12, one potential explanation is the type of instruction children were exposed to and Ethiopian children's generally low level of basic academic skills. While the PPVT, a basic measure of receptive vocabulary, is thought to be sensitive to a child's exposure to declarative knowledge about words, their definitions, and their usage (Nagy & Scott, 2000), it is relatively less sensitive to the national curriculum or to any particular pedagogical approach than math or reading comprehension tests, which contain more advanced skills and curriculum-sensitive content. Moreover, about a quarter of the 12-year-old Young Lives children (Younger Cohort) in Ethiopia could not read sentence(s) in 2013, and more than a third of 15-year-olds were unable to answer any of the math questions correctly in 2016 (Young Lives, 2017). My supplementary analysis also indicated that the preschool benefits were pronounced only in tasks in the mother tongue (e.g., oral reading fluency) that were easier than the reading comprehension tasks. Collectively, the fadeout of initial academic benefits may stem from the dissonance among the basic academic skill levels of Ethiopian children, their exposure to the right level of instruction, and the difficulty of tests.

Second, as for educational attainment, the results of the present study showed that preschool attendance led to significant increases in grade completion: by age 15, preschool attendees had accumulated about 0.7 additional years of schooling and were 25 percent more likely to progress adequately to age-appropriate grades. These results are in line with the previous work showing the benefits of ECE participation for higher educational attainment in LMICs (Berlinski et al., 2008; Bietenbeck et al., 2017; Hazarika & Viren, 2013; Krafft, 2015). In particular, despite the highly diverse LMIC settings, the magnitude of gains in the present study are similar to prior evidence that found a preschool effect of an additional 0.8 years of schooling at age 15 in Uruguay (Berlinski et al., 2008) and an additional year of schooling in Egypt

(Krafft, 2015). On the probability of achieving on-time grade progression, my findings are similar to Woldehanna and Araya's (2017) estimates using the Young Lives' Older Cohort: preschoolers were 25.7 percent more likely to complete secondary education at the proper age than their non-preschool peers.

Additionally, it is noteworthy that the benefits of preschool for later educational attainment do not exhibit 'convergence' or even an increase over students' educational trajectories. While the present study showed sustained preschool benefits in grade completion without a decline between ages 12 and 15, the results of studies in Uruguay (Berlinski et al., 2008), rural India (Hazarika & Viren, 2013), and Kenya and Tanzania (Bietenbeck et al., 2017) consistently point out that 'positive effects of preschool on school enrolment […] grows monotonically with age' (Berlinski et al., 2008, p. 1425). Further research needs to investigate this sustained or growing pattern that may support the multiplier effects of early learning inputs on future educational attainment (Cunha & Heckman, 2007), which present patterns that are distinct from those in academic benefits (e.g., IQ or test scores).

Third, the present study aimed to understand whether students respond differently to preschool attendance according to gender, household wealth, father's education level, and prior academic achievement. In terms of gender, the benefits of preschool tend to be larger among boys than among girls, but the differential influence as a function of gender is never statistically significant. These findings of no gender difference are similar to those from studies in Kenya and Tanzania (Bietenbeck et al., 2017), but only a few studies have assessed the differential effect of ECE by gender and the findings remain inconclusive (Barnett, 1995).

Notably, my findings based on household wealth and father's education level countered prior evidence showing particularly pronounced benefits of preschool for children from low-income households or those who live with poorly educated parents (U.K., App et al., 2013; Norway, Havnes & Mogstad, 2011; France, Dumas & Lefranc, 2012; Northern India, Hazarika & Viren, 2013; Uruguay, Berlinski et al., 2008). Unlike students from more affluent families, for whom the preschool benefits were sustained at age 15, the associations between preschool attendance and educational outcomes among students from poorer families are no longer evident in adolescence. It is possible to infer that the different life conditions of children with different socioeconomic backgrounds affect their ability to benefit from early learning. For children living in poverty, who are likely to grow up in less stimulating home learning environments, preschool participation alone may not be sufficient to overcome various barriers to and constraints on their ability to succeed throughout the life course (Brooks-gunn, 2003; Brooks-Gunn & Duncan, 1997). These results, which disappointingly are not aligned with ECE's compensatory prediction, imply that enriched early learning experiences can in fact amplify the learning inequality induced by socioeconomic disparities. This pattern appears to continue even after a large-scale expansion of public preschool, when children from a broad range of household wealth groups suddenly flooded into the pre-primary education system.

In contrast, there are relatively encouraging results from students' prior achievement levels. These results indicate that the association between preschool and child academic outcomes are greater among low achievers than high achievers. Contrary to these findings, prior evidence in the U.S. shows that the longer-term advantages of preschool were a function of children's earlier academic achievement (e.g., Ansari et al., 2017; Campbell et al., 2012), which is in line with predictions from complementary models. However, before concluding whether my findings are aligned with the compensatory models, there is a need for further investigation into the possibility of a ceiling effect among the high-achievement group and some irregular patterns observed within the middle-achievement group.

Fourth, the present study explored whether the association between preschool attendance and students' achievement at age 15 may differ by preschool characteristics. Several notable patterns did emerge: for academic achievement in PPVT and English at age 15, positive associations were observed among students who entered preschool earlier (at age 4) and attended private pre-primary institutions which provided high-quality teaching and full-time services.[154] These patterns are consistent with the prior evidence suggesting that preschool duration matters (e.g., Loeb et al., 2007, in the U.S.).[155] Within the LMIC context in particular, entrance to preschool before age 4 or attending preschool for at least two years were positively associated with improved academic achievement at age 11-12 in rural India (Singh & Makherjee, 2018) and Uganda (Hungi & Ngware, 2017), and with a higher probability of

[154] Note that children in the control group generally stayed at home, as there are few alternative options in Ethiopia.
[155] However, there are some mixed findings in the earlier studies by Barnett (1995) and Currie (2001) reporting that there is no significant effect, depending on age of entry to the U.S.

completing secondary education in Ethiopia (Woldehanna & Araya, 2017).[156] As for preschool type, provided that about four-fifths of preschools in Ethiopia were privately run during the period the data were collected, it may not be proper to generalise my finding that favours private preschool for students' academic achievement; however, attending a public preschool was highly associated with better educational attainment, which is equivalent to almost one additional year of schooling.

Further, the results of this study are aligned with the argument that preschool quality matters. Although little is known about the relation between preschool quality and child outcomes in low-resource settings, due to the lack of valid measures, my findings are consistent with earlier evidence from Bangladesh suggesting a positive association between quality preschool environments and children's first-grade performance (Aboud, 2006; Aboud et al., 2008). In line with my hypothesis that increased daily exposure to formal early learning environments may be beneficial for students' educational outcomes, the results of this study show that children attending preschool for more than seven hours a day appeared to have benefited more than those who attended fewer than seven hours. Based on these findings, more work is needed to understand various preschool characteristics that can influence the fundamental skills linked to individuals' long-term success, including age-appropriate curricula, teacher-child interactions, and teacher professional development. Moreover, as prior evidence points out that the positive associations between the daily preschool exposure and improved academic achievement are conditional on family income and race (Loeb et al., 2007), future research needs to explore the intersectionality between such complex processes as ECE programme quality, duration, and intensity and the characteristics of family and child for preserving the benefits of preschool across a variety of contexts, particularly in low-resource settings.

Fifth, the present study tested whether subsequent school environments could mediate the link between preschool attendance and student outcomes in adolescence. The school characteristics of upper primary grades that were investigated in the current analysis—school asset index, textbook-pupil ratio, low attendance, and teacher incentives—partially mediated the relation between preschool and students' PPVT scores at ages 12 and 15 when I treated these as the

[156] The age of preschool entry and duration are not necessarily consistent (e.g., early entry to preschool, then stop or dropout from preschool during the course), however, in this study, I assumed that early preschool entry (ages 4, 5, 6) leads to the longer duration (3, 2, 1 years) of preschool.

latent variable. This latent variable is similar to a composite score of four indicators related to school environments. In particular, the indirect path via subsequent school experience accounted for about one-third of the total association between preschool attendance and student's receptive vocabulary skills at ages 12 and 15. These findings are consistent with prior evidence from the U.S. from Curenton et al. (2015) and Reynolds et al. (2004); however, it is not possible to find commensurate evidence in LMICs, as few longitudinal studies of ECE programmes collected data on later school experiences. Determining the extent to which subsequent educational contexts facilitate the persistence of preschool influence has an important policy implication that can help to foster and sustain early academic advantages (Magnuson et al., 2007; Ramey & Ramey, 1998). Taken together, despite the fact that a number of theoretically driven preschool and primary school characteristics in the present study were relatively coarse and unlikely to capture the full scope of possible underlying mechanisms, this study highlights the importance of the institutional environments which children face at each stage of life that have the potential to promote or hinder the sustained benefits of ECE.

Lastly, the Young Lives sample (Younger Cohort) used in this chapter could be compatible with the EGRA 2010 sample used in the previous two chapters in terms of the ECE landscape in Ethiopia between 2005 and 2008.[157] However, according to the analysis using the most similar test in ORF measured at age 8, the opposite pattern is observed between the two samples: there was no significant association between preschool and ORF scores in the EGRA 2010 sample, whereas it became positive and statistically significant among the Young Lives sample (0.35 SD, $p < 0.01$). As noted earlier, this is attributed in part to the different sampling frame: while the EGRA 2010 was a regionally representative sample from the five regions in Ethiopia, the Young Lives Study oversampled the disadvantaged population purposively through a multi-stage sampling processes (Outes-Leon, 2011). To illustrate, the Young Lives sample showed a huge disparity in preschool access between urban (25.2%) and rural (2.6%) children, whereas the EGRA sample showed less disparity between urban (24%) and rural (11.3%) children. Meanwhile, both samples consistently showed no observable benefits of preschool on reading comprehension achievement.

[157] This is a period when both samples of children were preschool-eligible age.

4.9 Limitations

Although the Young Lives data provide a unique opportunity to assess the evolution of the relationship between preschool and child outcomes over time, there are several limitations that should be noted when interpreting results. First, these data captured the period when attending a preschool was regarded as a luxury that served a mere 5 percent of children in urban areas in Ethiopia out of seven million children. Hence, the Young Lives children attending preschool were distinct from the national Ethiopian population. This raises an issue of external validity, which limits the generalizability of my findings to the broader population, and to those who attended (or are attending) preschool in recent years. Additionally, this data limitation should be considered when interpreting the differential benefits of preschool between advantaged and disadvantaged children. There is a possibility that the observed differential influence by household wealth and parental education could be driven by differential selection—for example, with higher-income children over-represented in the sample attending preschool, especially in private and full-time preschool institutions. Although I adjusted for several non-random sources of preschool assignment, unobserved differences across these groups are likely to remain.

Second, despite covering a number of outcome measures for cognitive skills, this study was unable to explore non-cognitive skills—self-regulation, motivation, and persistence—that may contribute to the benefits of preschool for individuals' long-term success, according to prior evidence (Heckman et al., 2013).[158] Future research is needed that includes a broader range of cognitive and non-cognitive skills to measure the contribution of preschool attendance on various child developmental domains in both the short and long run. Third, as noted above, a lack of detailed data on preschool characteristics and subsequent school experience limits the conclusions that can be drawn in this study about mechanisms and implications for sustained gains from preschool. More work on both structural and process quality and their interaction with child and family characteristics will be necessary to identify preschool quality features that optimise student outcomes, particularly for low-resource, low-capacity settings.

[158] I also estimated the association between preschool attendance and non-cognitive skills available in the Young Lives data—self-efficacy, self-esteem, relationship with parents and peers—however, it never become significant (the results are available upon request).

Fourth, assuming that the impact of the unobservable time invariant is small, the current study used propensity score matching with a rich set of child and family covariates to yield a robust estimate that rules out many alternative explanations on differential selection into preschool. It is also notable that the propensity score models did little beyond the OLS regression models that account for 40-45 covariates, which perhaps implies that estimates from this methodology are likely close to the causal identification. In the meantime, interpretation needs to consider that it is not possible to completely rule out differential selection into preschool, particularly in LMICs, where household resources are limited and unpredictable (McCoy et al., 2017). In fact, households in Ethiopia frequently experience external shocks such as drought, flood, crop failure and ethnic clashes that negatively affect children's schooling (Berhane, Abay, & Woldehanna, 2016). Additionally, in the pursuit of long-term associations in the present study, selectivity occurred not only at the pre-primary level but continued at the primary and secondary levels. For example, parents who enrolled their children in preschool may in turn select higher quality primary schools. Lastly, with regard to the importance of considering the counterfactual's circumstances (Zhai et al., 2014), children in the control group likely stayed at home (non-formal home-based care) in the absence of alternative options in Ethiopia. However, there is a still possibility that they were involved in informal child care programmes or community-based social protection services, and in this case estimates may be biased downward.

4.10 Conclusion

Given the importance of early childhood development for outcomes later in life, investments in early childhood education are often considered promising investments with long-term pay-offs. The present study provides an important first step in understanding how the relation between preschool attendance and educational outcomes evolves over time in the context of a low-income country. The results of this study suggest that preschool attendance led to significant improvement in academic achievement and increased educational attainment of Ethiopian children at ages 8, 12, and 15. Nevertheless, some alarming patterns emerged in the differential benefits by household wealth, contrary to the common belief that an enriched early learning experience plays a role in 'equalizing' existing socioeconomic gradients in learning. In addition, des pite its limited ability to reflect the latest ECE landscape in Ethiopia, the present study sheds light on the importance of quality in preschool and subsequent school experiences. Notably, these quality dimensions have the potential not only to determine the preservation of preschool benefits but to facilitate students' positive academic trajectories from early childhood through adolescence.

In recent years, Ethiopia has been striving to transform the country's pre-primary education system through a rapid large-scale expansion of public preschool. However, this endeavour must not come at the expense of *equitable access to high-quality education* for all. While maintaining quality in the process of scaling up is a major challenge for every programme (Engle et al., 2011), particular attention must be paid in that expanding coverage to pursuing a systematic approach to the scale-up of ECE, including careful planning, targeted resourcing, and the continued monitoring and capacity-building of stakeholders as core strategies for reducing disparities and boosting the learning of Ethiopian children.

5 CHAPTER 5 – REFLECTION

Ethiopia's pre-primary education policy reform is continuously evolving. By exploiting the inception phase of reform, this dissertation found that expanded access *strengthens* the role preschool plays in predicting students' early grade reading outcomes; however, preschool expansion comes at the expense of equitable gains between advantaged and disadvantaged children. In that each of the previous chapters discusses my findings extensively, as well as their implications and limitations, and provides brief conclusions, this chapter offers my reflections on my research and field experiences in Ethiopia and on lessons learned for research and policy.

Since 2017, as part of the World Bank's education team, I've been privileged to observe and be involved in the ongoing early learning reform in Ethiopia. This reflection is based on my numerous notes from field visits and from meetings/interviews with policymakers and stakeholders in Ethiopia.[159] The reform has been an interplay between barriers to change and drivers for change; I focus on its potential and further room for improvement. I frame my reflection within three directions proposed by Yoshikawa and Nieto (2013) with respect to the emerging paradigm shifts for early childhood research, practice, and policy. These include (1) from impacts on the mean level of children's development to impacts on equity and inequality; (2) from quality to effectiveness factors in ECE settings, networks, and systems; and (3) from sectoral and multisectoral to community-based and participatory.

5.1 From Impacts on the Mean Level of Children's Development to Impacts on Equity and Inequality

The primary goal of the early learning reform in Ethiopia is to increase equity in the education system by reaching children from disadvantaged backgrounds, who are most at risk of exclusion, drop-out and under-achievement (MoE, 2015). My findings suggest that the rapid expansion of O-Class may not be achieving this goal and may instead amplify learning inequalities between rich and poor students. I examine not only the influences of ECE on average learning outcomes, but also the distribution of learning outcomes and other factors in a particular population. As national-level enrolment growth hides large discrepancies in

[159] The qualitative information is used with permission of the World Bank.

preschool access by location and household wealth, the average student achievement level masks large disparities in learning gains associated with preschool, particularly that the most disadvantaged children are least likely to benefit from pre-primary education. This reinforces the indication that an unsystematic approach to scale-up has not benefited all children equally.

Although O-Classes have started to reach many remote rural areas in Ethiopia, this singular effort to increase enrolment in preschool is not enough to reduce learning inequality among young children. Strategies that aim to achieve equity goals should go beyond the 'one-size-fits-all' policy approach. Yoshigawa and Nieto (2013) noted that 'closely monitoring ECD indicators in different subpopulations and using information with a framework of corrective and distributive justice can be an effective tool against unintended effects of ECD programmes in the direction of greater inequality or social exclusion' (p. 488). By regular monitoring of the equity indicators from various data sources available in Ethiopia (e.g., national statistics, national learning assessment, and household surveys), more proactive compensatory measures in ECE policy should be introduced to meet the needs of marginalised communities. This will enable all children to benefit and learn from enriched early learning experiences regardless of their geographic constraints or household economic status.

Moreover, about half of Ethiopian children still remain out of ECE. This alarming figure calls for giving more attention to reaching children currently deprived of early learning opportunities. There are two points to consider in recalibrating ECE policy and practice. First, targeted financial support should be provided based on resource levels and capacity for delivering ECE at the school and community level. The pitfall of the enrolment-based school grants for O-Class, which are the major financing source for ECE, is that they provide funding only for established O-Classes, not for schools without an O-Class.[160] Another financing source for O-Class is the block grants allocated to *woreda* (districts); however, due to competition with other education levels or other sectors within a woreda, little funding is allocated to pre-primary education.[161] The current approaches—a uniform funding allocation

[160] School grants (an enrolment-based capitation grant provided directly to school) for primary schools have been provided since 2013 under the GEQIP-II programme. In 2016-2017, school grants for O-Class were introduced as a top-up to the regular school grants for primary in order to reflect growing demand for pre-primary education.
[161] From personal interviews with regional and local stakeholders.

formula and no grants earmarked for ECE—have the potential to reinforce the vicious cycle of ‘the rich get richer and the poor get poorer’ that leave behind the schools most in need.

Second, targeted interventions should be provided for those without ECE experience, but in an inclusive manner. In 2015, as one potential way to intervene, Ethiopia introduced the Accelerated School Readiness programme, a short-term supplementary course during summer break or the first two months of Grade 1. During my school visits in Ethiopia, I encountered settings similar to those of my research design—the Grade 1 classroom using *ability grouping* according to students who attended kindergarten, those attended O-Class, and those who did not attend either. Although observed differences in reading ability across the three groups reaffirmed the importance of early learning, more concerns were raised about how this ability grouping approach affects child development.[162] Labelling a child as ‘unprepared’ at school entry may have a negative influence on their development, especially on their socio-emotional skills such as self-esteem, motivation, and persistence. Therefore, any interventions that aim to remedy early learning loss should be academically motivated and emotionally sensitive to the targeted group.

5.2 From Quality to Effectiveness Factors in ECE Settings, Networks, and Systems

A rapid expansion of preschool comes with the risk of lowering quality; however, the government’s focus is gradually shifting to improving quality. This shift involves the introduction of a practice-based, interactive teacher training for O-Class teachers in two educationally disadvantaged regions, Benishangul-Gumuz and Gambella, which were implemented in 2017-2018.[163] This pilot intervention, led by the government with technical and financial support from the World Bank, includes the development and implementation of in-service teacher training and the provision of curriculum materials, including some basic play materials, to O-Classes. The training consists of comprehensive knowledge of child development, interactive pedagogical approaches, and hands-on activities to facilitate play-based learning. More than 700 teachers were trained, which exceeded the initial target for the

[162] Note that this is an anecdotal case that cannot be generalised to practices in other schools in Ethiopia. International evidence also reached a consensus that ability grouping has a negative impact, especially for those who struggle most with their learning (McGillicuddy & Devine, 2017).
[163] This pilot programme is part of the World Bank’s Ethiopia Education Results Based Financing project financed by the Global Partnership for Education.

two regions, and demand from the local communities continues to grow. Training participants reported that the new training was a ‘transformative’ experience which was very distinct from traditional training, and that it in particular identified (i) perception changes in ECE; (ii) improved confidence in activity-based learning and increased teacher-child interactivity; (iii) better understanding of more inclusive approaches for meeting the diverse needs of young children; (iv) learning skills for classroom management and lesson planning; and (v) creating locally adapted learning materials.[164]

Moreover, beyond these two promising quality elements, new teacher training and curriculum, the experience gained in the two regions benefited the pre-primary education sector in Ethiopia through capacity-building and horizontal and vertical collaboration among policy networks. As Hommel (2013) pointed out, while the traditional focus of a scale-up initiative is on training and capacity-building for frontline providers (i.e., teachers), the capacity of various stakeholders—federal and regional government officers, development partners, researchers, NGOs, and community and school leaders—may be just as critical in successful ECE systems.

The first advance is capacity-building among key stakeholders at different levels. To illustrate, at the central level, although the initial training was led by ECE experts from development partners and NGOs, the lead responsibility gradually shifted to experts at the Ministry of Education and an established talent pool from the earlier trainings; the recent scale-up initiatives have been fully managed by the federal government. At the regional level, the implementers’ leadership and institutional capacity have been strengthened. In the Benishangul-Gumuz region, Regional Education Bureau (REB) led the locally adapted teacher training by ensuring the involvement of teachers of local ethnicity and incorporating local culture and languages into activities. The effective management model used in this region, a technical ECE working group composed of a steering committee and a technical committee of regional officials, has been replicated in other regions.[165] After the teacher training, the REB organized and supported the ECE training for school leaders and supervisors by their own capacity. Capacity-building also has expanded to the Colleges of Teachers’ Education, as some

[164] Based on interviews with trainees and pre- and post-training questionnaires.
[165] In a technical working group, a steering committee is responsible for planning, monitoring, and financing; and a technical committee is responsible for management, logistic, and day-to-day/follow-up activities.

college instructors have applied their new knowledge and approaches to improve pre-service training by, for example, shifting from theory-oriented to play-oriented training.

The second advance is collaboration among key stakeholders. From the development of curriculum materials to the implementation and monitoring of the intervention, horizontal and vertical links have been strengthened across the MoE, REBs, developmental partners, and researchers (e.g., UNICEF, World Bank, Save the Children, Right to Play, and Kotebe Metropolitan University) in collaboration with teachers, schools, and communities. Successful implementation was made possible by the rise of 'policy networks' that share a common set of values and goals through new partnerships of political actors (Sabatier & Jenkins-Smith, 1993).

Building on these efforts, the national flagship reform programme, GEQIP-E, introduced the Quality Enhancement and Assurance Programme (QEAP) for O-Class in 2018 (World Bank, 2017).[166] QEAP aims to provide a comprehensive package of interventions to systemically improve the quality of O-Class provision within a coherent framework. To address the observed problem of 'working in silos', QEAP intends to provide an incentive for improved coordination across interventions and key stakeholders working on O-Class quality improvement, without which meaningful results at the classroom level cannot be achieved. The design of QEAP considers in particular the fact that each of a set of policy levers—teacher professional development, curriculum, quality monitoring, and data-based accountability—has its own properties and limitations but may 'act in interdependent fashion to determine the ultimate effectiveness of ECE policies in enhancing child development at scale' (Yoshigawa et al., 2018, p. 8). Although some uncertainty and challenges remain in finding answers to the questions, 'how can ECE systems be improved to deliver the results?' research using the systems approach is emerging and growing in the ECE field (Kagan & Kauerz, 2012). Future research needs to delve into the interlinked effective factors—the key dimensions of quality associated with more positive effects—that work together to meet the practice and policy needs of ECE systems.

[166] QEAP for O-Classes comprises two key components: Quality Enhancement (QE); and Quality Assurance (QA). The QE component, which improves pedagogical practices in the classroom, includes teacher preparation and professional development, curriculum, and TLM for O-Class, and training for management and supervision. The QA component establishes national standards, an inspection process, and EMIS data collection (World Bank, 2017, p. 41).

5.3 From Sectoral and Multisectoral to Community-Based and Participatory

The initial phase of the early learning reform was implemented in a top-down manner directed by the central governments. The downward delegation from federal and regional governments to woredas and schools often suffered from a lack of coordination and poor communication, which led to ‘incoherence between national objectives and local preferences and capacities’ (Rossiter et al., 2018). As each region forged ahead with its own plan for preschool expansion, there were both challenges and opportunities. In other words, the various expansion efforts resulted in huge regional disparities due to the different levels of financial and human resources; however, they also resulted in innovative service delivery tailored to the needs of local communities. For example, given that school proximity is a key factor in a child’s schooling outcomes, O-Classes in SNNP utilised the existing community-based organizations, such as religious institutions, farmers’ training centres, and established community centres. The Accelerated School Readiness programme in Afar was adapted to reflect the needs of pastoralist communities and their mobile school systems.

In the absence of a dedicated budget at the central level, community contributions have been a main driver in the expansion of preschool in Ethiopia. Although my dissertation focuses on ECE policies at the federal and regional levels, Rossiter et al., (2018) stressed that, with a strong sense of community ownership of O-Class, ‘community investments and collaboration in early learning service provision assure strong voice in negotiations of which services are going to be delivered’ (p. 30). In fact, during my school visits, I observed the significant contributions the parent-teacher associations and community leaders made to the provision of early learning, including the construction of classrooms and latrines for young children, the hiring of local teacher assistants, and the provision of basic stationery and/or school meals. With peer-to-peer support within and between communities, ‘upward’ delegation from community to school can be more flexible and responsive to a community’s need to integrate local cultural norms. Meanwhile, careful attention must be paid to variations in local beliefs and values about children and families, and to resource constraints, which often are perceived as barriers to ECE implementation.

Participatory processes, particularly those involving regional/local stakeholders, parents, community members, and children themselves, are an important link between community-based input and policy development. The government of Ethiopia is currently preparing a revision of the National Early Childhood Care and Education framework. In the revision process, strategies that aim to increase participation at the community level should go beyond representing the interests of the government. It should be a collaborative process that captures the expectations and demands, as well as innovative ideas from, the local and community stakeholders. A deep understanding of the complex ways policy initiatives intersect with local realities is needed to achieve this goal. Specifically, the new policy framework can support the regional stakeholders in developing region-tailored goals and strategies to promote local ownership of ECE provision and ameliorate sub-geographical inequalities in ECE. More investment of time and resources is needed to create a platform that reflects local voices and guarantees the participation of communities and local governments. These efforts will promote an open dialogue between local actors and national/regional policymakers.

A primary contribution of this dissertation is to inform future ECE policy in Ethiopia, and possibly in other LMICs. Further improvement in these areas—promoting equitable access to and gains from preschool, strengthening quality and institutional capacity in the ECE systems, and ensuring a participatory approach with stakeholders at all levels—requires both time and a political and financial commitment. It also calls for creative thinking about interconnected but under-researched issues, including the multi-aged O-Classes, the learning continuum of the foundational grades, a sustainable financing mechanism, the ECE workforce, public and private partnerships, and multi-sectoral governance. The positive momentum created by the ongoing early learning reform in Ethiopia must be continued in order to build up inclusive and coherent ECE systems.

REFERENCES

Abebe, W. (2012). *School Management and Decision-making in Ethiopian Government Schools: Evidence from the Young Lives Qualitative School Survey*. Working paper 86, Oxford: Young Lives.

Aboud, F. E. (2006). Evaluation of an early childhood preschool program in rural Bangladesh. *Early Childhood Research Quarterly*, *21*(1), 46–60.

Aboud, F. E., & Hossain, K. (2011). The impact of preprimary school on primary school achievement in Bangladesh. *Early Childhood Research Quarterly*, *26*(2), 237–246.

Aboud, F. E., Hossain, K., & O'Gara, C. (2008). The Succeed Project: Challenging Early School Failure in Bangladesh. *Research in Comparative and International Education*, *3*(3), 295–307.

Agirdag, O., Yazici, Z., & Sierens, S. (2015). Trends in pre-school enrolment in Turkey: unequal access and differential consequences. *Comparative Education*, *51*(4), 537–554.

Aguilar, R., & Tansini, R. (2012). Joint analysis of preschool attendance and school performance in the short and long-run. *International Journal of Educational Development*, *32*(2), 224–231.

Aizer, A., & Cunha, F. (2012). *The production of human capital: endowments, investments and fertility*. NBER Working Papers 18429, National Bureau of Economic Research, Inc, Cambridge, MA.

Alcott, B. (2017). Does Teacher Encouragement Influence Students' Educational Progress? A Propensity-Score Matching Analysis. *Research in Higher Education*, *58*(7), 773–804.

Alderman, H., & Vegas, E. (2011). *No Small Matter* (H. Alderman, ed.). Washington, D.C:World Bank.

Allison, P. D. (2002). *Missing Data*.Thousand Oaks, CA:Sage Publications.

Allison, P. D. (2009). *Fixed effects regression models*. Thousand Oaks, CA: Sage Publications.

Altinok, N. (2009). An empirical approach to marginalization in education based on the TIMSS 2007 study. Background paper for EFA Global Monitoring Report 2010. Paris: UNESCO.

American Institutes for Research. (2013). *Getting Ready for School. A Child-to-Child Approach - Program Evaluation Report*. Washington, D.C.: American Institutes for Research.

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. Washington, D.C.: American Institutes for Research.

Ansari, A. (2018). The persistence of preschool effects from Early Childhood through adolescence. *Journal of Educational Psychology*, *110*(7), 952–973.

Ansari, A., López, M., Manfra, L., Bleiker, C., Dinehart, L. H. B., Hartman, S. C., & Winsler, A. (2017). Differential Third-Grade Outcomes Associated With Attending Publicly Funded Preschool Programs for Low-Income Latino Children. *Child Development*, *88*(5), 1743–1756.

Apps, P., Mendolia, S., & Walker, I. (2013). The impact of pre-school on adolescents' outcomes: Evidence from a recent English cohort. *Economics of Education Review*, *37*, 183–199.

Araujo, M. C., Carneiro, P., Cruz-Aguayo, Y., & Schady, N. (2016). *A Helping Hand? Teacher Quality and Learning Outcomes in Kindergarten*. IDB Working Paper Series; 665. Inter-American Development Bank.

Armecin, G., Behrman, J., Duazo, P., Ghuman, S., Gultiano, S., King, E., et al. (2006). *Early childhood development through an integrated P\program: Evidence from the Philippines*. (Working Paper 3922). Washington, D.C. World Bank Policy Research.

Attanasio, O., Meghir, C., Nix, E., & Salvati, F. (2017). Human capital growth and poverty: Evidence from Ethiopia and Peru. *Review of Economic Dynamics*, *25*, 234–259.

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

Bailey, D., Duncan, G. J., Odgers, C. L., & Yu, W. (2017). Persistence and Fadeout in the Impacts of Child and Adolescent Interventions. *Journal of Research on Educational Effectiveness*, *10*(1), 7–39.

Baltes, P. B., Lindenberger, U., & Staudinger, U. M. (1998). Life-span theory in developmental psychology. In W. Damon & R. M. Lerner (Eds.), *Handbook of child psychology: Theoretical models of human development* (pp. 1029–1143). Hoboken, NJ, US: John Wiley & Sons Inc.

Banerjee, A., Banerji, R., Duflo, E., Glennerster, R., & Khemani, S. (2008). *Pitfalls of Participatory Programs*. NBER Working Papers 14311, National Bureau of Economic Research, Inc, Cambridge, MA.

Barnett, W. S. & Belfield, C. R. (2006). Early childhood development and social mobility. *The Future of Children*, *16*(2), 73–98.

Barnett, W. S. (1995). Long-Term Effects of Early Childhood Programs on Cognitive and School Outcomes. *The Future of Children*, *5*(3), 25.

Barnett, W. S. (2008). Preschool education and its lasting effects: Research and policy implications. *National Institute for Early Education Research*, 1–37.

Barnett, W. S., Jung, K., Youn, M.-J., & Frede, E. (2013). Abbott preschool program longitudinal effects study: Fifth grade follow-up. *National Institute for Early Education Research*, 1–34.

Baron, R., & Kenny, D. (1986). The moderator-mediator variable distinction in social psychological research: Conceptual, strategic, and statistical considerations. *Journal of personality and social psychology*, *51*(6), 1173.

Bartlett, L., Dowd, A. J., & Jonason, C. (2015). Problematizing early grade reading: Should the post-2015 agenda treasure what is measured? *International Journal of Educational Development*, *40*, 308–314.

Bartus, T. (2005). Estimation of marginal effects using margeff. *Stata Journal*, *5*(3), 309–329.

Bashir, S., Lockheed, M. E., Ninan, E., & Tan, J.-P. (2018). *Facing Forward: Schooling for Learning in Africa*. Washington, D.C.: World Bank.

Bassok, D., Gibbs, C. R., & Latham, S. (2019). Preschool and Children's Outcomes in Elementary School: Have Patterns Changed Nationwide Between 1998 and 2010? *Child Development*, *90*(6), 1875–1897.

Bastos, P., Bottan, N. L., & Cristia, J. (2017). Access to Preprimary Education and Progression in Primary School: Evidence from Rural Guatemala. *Economic Development and Cultural Change, 65*(3), 521-547.

Becker, G. S. (1962). Investment in Human Capital: A Theoretical Analysis. *Journal of Political Economy*, 70 (5, Part 2), 9-49.

Becker, G. S. (1967). *Human capital and the personal distribution of income*. Ann Arbor, MI: Institute of Public Administration.

Becker, G. S. (1975). Investment in human capital: effects on earnings. *In Human Capital: A Theoretical and Empirical Analysis, with Special Reference to Education,* Second Edition (pp. 13-44). National Bureau of Economic Research, Inc, Cambridge, MA.

Becker, S. O., & Caliendo, M. (2007). *mhbounds – Sensitivity Analysis for Average Treatment Effects.* Discussion Papers of DIW Berlin 659, DIW Berlin, German Institute for Economic Research.

Becker, S. O., & Ichino, A. (2002). Estimation of Average Treatment Effects Based on Propensity Scores. *The Stata Journal: Promoting Communications on Statistics and Stata*, *2*(4), 358–377.

Belfield, C. R., Nores, M., Barnett, S., & Schweinhart, L. (2006). The High/Scope Perry Preschool Program: Cost–Benefit Analysis Using Data from the Age-40 Followup. *Journal of Human Resources, XLI*(1), 162–190.

Bentler, P. M. (1990). Comparative fit indexes in structural models. *Psychological Bulletin*, *107*(2), 238–246.

Berhane, G., Abay, M. H., & Woldehanna, T. (2016). *Childhood shocks, safety nets and cognitive skills: Panel data evidence from rural Ethiopia*. Working Paper 73. Ethiopian Development Research Institute & International Food Policy Research Institute.

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.

Berlinski, S., Galiani, S., & Manacorda, M. (2008). Giving children a better start: Preschool attendance and school-age profiles. *Journal of Public Economics, 92*(5–6), 1416–1440.

Bharath, S. T., Dahiya, S., Saunders, A., & Srinivasan, A. (2011). Lending relationships and loan contract terms. *Review of Financial Studies*, *24*(4), 1141-1203.

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.

Black, M. M., Walker, S. P., Fernald, L. C. H., Andersen, C. T., DiGirolamo, A. M., Lu, C., … Grantham-McGregor, S. (2017). Early childhood development coming of age: science through the life course. *The Lancet, 389*(10064), 77–90.

Blair, C., & Raver, C. (2012). Child Development in the Context of Adversity. *American Psychologist, 67*(4), 309–318.

Blundell, R., Dearden, L., & Sianesi, B. (2005). Evaluating the Effect of Education on Earnings:

Models, Methods and Results from the. *Journal of the Royal Statistical Society. Series A (Statistics in Society)*, *168*(3), 473–512.

Bogard, K., & Takanishi, R. (2005). PK-3: An Aligned and Coordinated Approach to Education for Children 3 to 8 Years Old. *Social Policy Report*, *19*(3), 1–24.

Bold, T., Kimenyi, M., Mwabu, G., Ng'ang'a, A., & Sandefur, J. (2018). Experimental evidence on scaling up education reforms in Kenya. *Journal of Public Economics*, *168*, 1–20.

Bond, T. N., & Lang, K. (2013). The evolution of the black-white test score gap in grades K-3: The fragility of results. *Review of Economics and Statistics*, *95*(5), 1468–1479.

Bouguen, A., Filmer, D., Macours, K., & Naudeau, S. (2014). *Preschool and parental response in a second best world: Evidence from a school construction experiment.* PSE Working Papers N° 2014-35.

Brambor, T., Clark, W. R., & Golder, M. (2006). Understanding Interaction Models: Improving Empirical Analyses. *Political Analysis*, *14*(1), 63–82.

Brinkman, S. A., Hasan, A., Jung, H., Kinnell, A., & Pradhan, M. (2017). The Impact of Expanding Access to Early Childhood Education Services in Rural Indonesia. *Journal of Labor Economics*, *35*(S1), S305–S335.

Britto, P. R., Yoshikawa, H., & Boller, K. (2011). Quality of Early Childhood Development Programs in Global Contexts: Rationale for Investment, Conceptual Framework and Implications for Equity. Social Policy Report. Volume 25, Number 2. *Society for Research in Child Development*, *25*(2).

Britton, J., & Vignoles, A. (2017). Education production functions. In G. Johnes, J. Johnes, T. Agasisti, & L. Lopez-Torres (Eds.), *Handbook of Contemporary Education Economics*. Edward Elgar M.U.A.

Bronfenbrenner, U., & Morris, P. A. (2006). The Bioecological Model of Human Development. In I. R. M. Lerner & W. Damon (Eds.), *Handbook of child psychology: Theoretical models of human development* (pp. 793–828). Hoboken, NJ, US: John Wiley & Sons Inc.

Bronfenbrenner, U. (1979). *Ecology of human development: Experiments by nature and design*. Cambriege, MA: Havard University Press.

Bronfenbrenner, U. (1986). Ecology of the Family as a Context for Human Development. Research Perspectives. *Developmental Psychology*, *22*, 723-742.

Brooks-gunn, J. (2003). Do You Believe In Magic?: What We Can Expect From Early Childhood Intervention Programs. *Social Policy Report*, *XVII*(1).

Brooks-Gunn, J., & Duncan, G. J. (1997). The Effects of Poverty on Children. *The Future of Children*, *7*(2), 55–71.

Bryant, D.M., Zaslow, M., & Burchinal,M. (2010). Issues inmeasuring program quality. In P. W. Wesley, & V. Buysse (Eds.), *The quest for quality: Promising innovations for early childhood programs* (pp. 47–68). Baltimore, MD: Brookes Pub. Co.

Burger, K. (2010). How does early childhood care and education affect cognitive development? An international review of the effects of early interventions for children from different social backgrounds. *Early Childhood Research Quarterly, 25*(2), 140–165.

Caliendo, M., & Kopenig, S. (2008). Some practical guidance for the implementation of propensity score matching. *Journal of Economic Surveys, 22*(1), 31–72.

Cameron, A. C., & Trivedi, P. K. (2010). *Microeconometrics Using Stata, revised edition.* StataCorp LP.

Campbell, F. A., Pungello, E. P., Burchinal, M., Kainz, K., Pan, Y., Wasik, B. H., … Ramey, C. T. (2012). Adult outcomes as a function of an early childhood educational program: An Abecedarian Project follow-up. *Developmental Psychology*, *48*(4), 1033–1043.

Campbell, F. A., Ramey, C. T., Pungello, E., Sparling, J., & Miller-Johnson, S. (2002). Early Childhood Education: Young Adult Outcomes From the Abecedarian Project. *Applied Developmental Science*, *6*(1), 42–57.

Campbell, F., Conti, G., Heckman, J. J., Moon, S. H., Pinto, R., Pungello, E., & Pan, Y. (2014). Early Childhood Investments Substantially Boost Adult Health. *Science*, *343*(6178), 1478–1485.

Carneiro, P., & Heckman, J. (2003). *Human capital policy*. NBER Working Papers 9495, National Bureau of Economic Research, Inc, Cambridge, MA.

Cascio, E. (2009). *Do Investments in Universal Early Education Pay Off? Long-term Effects of Introducing Kindergartens into Public Schools*. NBER Working Papers 14951, National Bureau of Economic Research, Inc, Cambridge, MA.

Central Statistical Agency, & World Bank. (2017). *LSMS-Integrated Surveys on Agriculture Ethiopia Socioeconomic Survey (ESS) 2015/2016*. Addis Ababa, Central Statistical Agency.

Central Statistical Agency (CSA) [Ethiopia] and ICF. 2016. *Ethiopia Demographic and Health Survey 2016*. Addis Ababa, Ethiopia, and Rockville, Maryland, USA: CSA and ICF.

Chetty, R., Friedman, J. N., Hilger, N., Saez, E., Schanzenbach, D. W., & Yagan, D. (2011). How does your kindergarten classroom affect your earnings? Evidence from project star. *Quarterly Journal of Economics*, *126*(4), 1593–1660.

Claessens, A., Engel, M., & Curran, F. C. (2014). Academic Content, Student Learning, and the Persistence of Preschool Effects. *American Educational Research Journal, 51*(2), 403–434.

Clarke, P., Crawford, C., Steele, F., & Vignoles, A. (2015). Revisiting fixed-and random-effects models: some considerations for policy-relevant education research. *Education Economics, 23*(3), 259–277.

Cobb, C. W., & Douglas, P. H. (1928). A Theory of Production. *American Economic Review, 18*(1), 139–165.

Cohen, J. (1992). A power primer. *Psychological bulletin, 112*(1), 155.

Conn, K. (2014). *Identifying Effective Education Interventions in Sub-Saharan Africa: A meta-analysis of rigorous impact evaluations.* Doctoral dissertation. New York: Teachers College, Columbia University.

Connor, M. O., Gray, S., Tarasuik, J., Connor, E. O., Kvalsvig, A., Incledon, E., & Goldfeld, S. (2016).

Early Childhood Research Quarterly Preschool attendance trends in Australia : Evidence from two sequential population cohorts. *Early Childhood Research Quarterly*, *35*, 31–39.

Crouch, L., & Destefano, J. (2017). *Doing Reform Differently: Combining Rigor and Practicality in Implementation and Evaluation of System Reforms*. International development group working paper no. 2017-01, RTI International, Research Triangle Park, NC.

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.

Crump, R. K., Hotz, V. J., Imbens, G. W., & Mitnik, O. A. (2009). Dealing with limited overlap in estimation of average treatment effects. *Biometrika*, *96*(1), 187–199.

Cueto, S., & Leon, J. (2012). *Psychometric Characteristics of Cognitive Development and Achievement Instruments in Round 3 of Young Lives.*Technical Paper 25: Oxford, Young Lives.

Cunha, F., & Heckman, J. (2007). The technology of skill formation. *American Economic Review*, *97*(2), 31–47.

Cunha, F., Heckman, J. J., Lochner, L., & Masterov, D. V. (2006). Chapter 12 Interpreting the Evidence on Life Cycle Skill Formation. *Handbook of the Economics of Education*, *1*(06), 697–812.

Curenton, S. M., Dong, N., & Shen, X. (2015). Does aggregate school-wide achievement mediate fifth grade outcomes for former early childhood education participants? *Developmental Psychology*, *51*(7), 921–934.

Currie, J. (2001). Early Childhood Education Programs. *Journal of Economic Perspectives*, *15*(2), 213–238.

Currie, J. (2009). Healthy, Wealthy, and Wise: Socioeconomic Status, Poor Health in Childhood, and Human Capital Development. *Journal of Economic Literature*, *47*(1), 87–122.

Currie, J., & Thomas, D. (1995). Does Head Start Make a Difference ? *The American Economic Review*, *85*(3), 341–364.

Currie, J., & Thomas, D. (1999). Does Head Start help hispanic children? *Journal of Public Economics*, *74*(2), 235–262.

Currie, J., & Thomas, D. (2000). School Quality and the Longer-Term Effects of Head Start. *Journal of Human Resources*, *35*(4).

Davis-Kean, P. E. (2005). The Influence of Parent Education and Family Income on Child Achievement: The Indirect Role of Parental Expectations and the Home Environment. *Journal of Family Psychology*, *19*(2), 294–304.

Deming, D. (2009). Early Childhood Intervention and Life-Cycle Skill Development: Evidence from Head Start. *American Economic Journal: Applied Economics*, *1*(3), 111–134.

DiPrete, T. A., & Gangl, M. (2004). Assessing Bias in the Estimation of Causal Effects: Rosenbaum Bounds on Matching Estimators and Instrumental Variables Estimation with Imperfect Instruments. *Sociological Methodology, 34*(2004), 271–310.

DiPrete, T. A., & Jennings, J. L. (2012). Social and behavioral skills and the gender gap in early educational achievement. *Social Science Research*, *41*(1), 1–15.

Dom, C. (2010). *2009. Desk Study: Ethiopia. Mid-term Evaluation of the EFA Fast Track Initiative.* Discussion Draft. London: Mokoro Consultants.

Dowd, A. J., Borisova, I., Amente, A., & Yenew, A. (2016). Realizing Capabilities in Ethiopia: Maximizing Early Childhood Investment for Impact and Equity. *Journal of Human Development and Capabilities*, *17*(4), 477–493.

Dubeck, M. M., & Gove, A. (2015). The early grade reading assessment (EGRA): Its theoretical foundation, purpose, and limitations. *International Journal of Educational Development*, *40*, 315–322.

Duflo, E., Dupas, P., & Kremer, M. (2015). School governance, teacher incentives, and pupil–teacher ratios: Experimental evidence from Kenyan primary schools. *Journal of Public Economics, 123,* 92-110.

Dumas, C., & Lefranc, A. (2010). Early schooling and later outcomes : Evidence from preschool extension in France, in Ermisch. *Université de Cergy-Pontoise, THEMA*, (81).

Duncan, G J, Brooks-Gunn, J., & Klebanov, P. K. (1994). Economic deprivation and early childhood development. *Child Development*, *65*(2), 296–318.

Duncan, Greg J, & Magnuson, K. (2013). Investing in Preschool Programs. *Journal of Economic Perspectives*, *27*(2), 109–132.

Dunn, L. M., & Dunn, L. M. (1997). *Peabody Picture Vocabulary Test*—Third Edition. Circle Pines, MN: American Guidance Service.

Engel, J. (2011). *Ethiopia's progress in education: A rapid and equitable expansion of access*. London: Oversea Development Institute.

Engle, P. L., Fernald, L. C. H., Alderman, H., Behrman, J., O'Gara, C., Yousafzai, A., … 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.

Entwisle, D. R., Alexander, K. L., & Olson, L. S. (2007). Early Schooling: The Handicap of Being Poor and Male. *Sociology of Education*, *80*(2), 114–138.

Favara, M. (2016). *Do dreams come true? Aspirations and educational attainments of Ethiopian boys and girls*. Working Paper 146: Oxford, Young Lives.

Feinstein, Leon. (2003). inequality in the early cognitive development of british children in the 1970 cohort. *Economica*, *70*, 73–97.

Fitzpatrick, M. D. (2010). Preschoolers Enrolled and Mothers at Work? The Effects of Universal Pre-Kindergarten. *Journal of Labor Economics, 28(1), 51-85.*.

Ford, D. H., & Lerner, R. M. (1992). *Developmental systems theory: An integrative approach.* Thousand Oaks, CA: Sage.

Frolich, M. (2004). Programme Evaluation With multiple treatments. *Journal of Economic Surveys, 18*(2), 181–224.

Frölich, M., & Michaelowa, K. (2011). Peer effects and textbooks in African primary education. *Labour Economics*, *18*(4), 474–486.

Fuchs, L. S., Fuchs, D., Hosp, M. K., & Jenkins, J. R. (2001). Oral Reading Fluency as an Indicator of Reading Competence: A Theoretical, Empirical, and Historical Analysis. *Scientific Studies of Reading*, *5*(3), 239–256.

Garces, E., Thomas, D., & Currie, J. (2002). Longer-term effects of head start. *American Economic Review*, *92*(4), 999–1012.

Garson, G. D. (2015). Missing values analysis and data imputation. *Statistical Associates Publishers*, 1–26.

Gasper, J., DeLuca, S., & Estacion, A. (2012). Switching Schools: Revisiting the Relationship Between School Mobility and High School Dropout. *American Educational Research Journal*, *49*(3), 487–519.

Gertler, P., Heckman, J., Pinto, R., Zanolini, A., Vermeerch, C., Walker, S., … Grantham-McGregor, S. (2014). Labor Market Returns to an Early Childhood Stimulation Intervention in Jamaica. *Science*, *344*(6187), 998–1001.

Gibbs, C., Ludwig, J., & Miller, D. (2011). Does Head Start do any lasting good? NBER Working Papers 17452, National Bureau of Economic Research, Inc, Cambridge, MA.

Glewwe, P., Kremer, M., & Moulin, S. (2009). Many children left behind? Textbooks and test scores in Kenya. *American Economic Journal: Applied Economics*, *1*(1), 112-35.

Glewwe, P. W., Hanushek, E. A., Humpage, S. D., & Ravina, R. (2013). School resources and educational outcomes in developing countries: A review of the literature from 1990 to 2010. *Education Policy in Developing Countries*, 13–64.

Goldstein, H. (1995). Hierarchical Data Modeling in the Social Sciences. *Journal of Educational and Behavioral Statistics*, *20*(2), 201–204.

Good, R.H., & Kaminski, R. A. (Eds.) (2002). *Dynamic Indicators of Basic Early Literacy Skills* (6th ed.). Eugene, OR: Institute for the Development of Educational Achievement.

Goodman, A., & Sianesi, B. (2005). Early education and children's outcomes: How long do the impacts last? *Fiscal Studies*, *26*(4), 513–548.

Gormley Jr., W. T. (2008). The Effects of Oklahoma's Pre-K Program on Hispanic Children. *Social Science Quarterly*, *89*(4), 916–936.

Gormley, W. T., Gayer, T., Phillips, D., & Dawson, B. (2005). The effects of universal pre-k on cognitive development. *Developmental Psychology*, *41*(6), 872–884.

Gottlieb, G., Wahlsten, D., & Lickliter, R. (2007). The significance of biology for human development: A developmental psychobiological systems view. *Handbook of child psychology, 1*.

Gove, A., Brunette, T., Bulat, J., Carrol, B., Henny, C., Macon, W., … Sitabkhan, Y. (2017). Assessing the impact of early learning programs in Africa. *New directions for child and adolescent development*, *2017*(158), 25-41.

Gove, A., & Cvelich, P. (2011). *Early reading: Igniting education for all. A report by the early grade learning community of practice*. North Carolina: RTI International.

Gove, A., Dombrowski, E., Pressley, J., King, S., & Weatherholt, T. (2018). Persistence and fadeout of preschool participation effects on early reading skills in low and middle-income countries. *Global Education Review*, *5*(2), 85–109.

Gove, A., & Wetterberg, A. (2011). *The Early Grade Reading Assessment: Applications and Interventions to Improve Basic Literacy.* North Carolina: RTI International.

Government of Ethiopia. (1994). *Education and Training Policy of the Federal Democratic Republic Government of Ethiopia*. Addis Ababa: Government of Ethiopia.

Grantham-McGregor, S., Cheung, Y. B., Cueto, S., Glewwe, P., Richter, L., & Strupp, B. (2007). Developmental potential in the first 5 years for children in developing countries. *Lancet*, *369*(9555), 60–70.

Guo, S., & Fraser, M. W. (2015). *Propensity score analysis : statistical methods and applications*. Second Edition., Thousand Oaks, Sage Publications, Inc.

Hanushek, E. A. (2002). *Publicly provided education*. NBER Working Papers 8799, National Bureau of Economic Research, Inc, Cambridge, MA.

Hanushek, E. A. (2003). The failure of input-based schooling policies. *Economic Journal*, *113*(485), 64–98.

Haveman, R., & Wolfe, B. (1995). The determinants of chidlren's attainments: A review of methods and findings. *Journal of Economic Literature*, *33*(4), 1829–1878.

Havnes, T., & Mogstad, M. (2011). No Child Left Behind: Universal Child Care and Children's Long-Run Outcomes. *American Economic Journal: Economic Policy*, *3*(May), 97–129.

Hazarika, G., & Viren, V. (2013). The effect of early childhood developmental program attendance on future school enrollment and grade progression in Rural North India. *Economics of Education Review*, *34*, 146–161.

Heckman, J. J. (2006). Skill formation and the economics of investing in disadvantaged children. *Social Sciences*, *312*(June), 2005–2007.

Heckman, J. J. (2011). The Economics of Inequality. The Value of Early Childhood Education. *American Educator*, 31–36.

Heckman, J. J., Ichimura, H., & Todd, P. E. (1998). Matching As An Econometric Evaluation Estimator. *Review of Economic Studies*, *65*, 261–294.

Heckman, J. J., Malofeeva, L., Pinto, R., & Savelyev, P. (2010). Understanding the Mechanisms Through which an Influential Early Childhood Program Boosted Adult Outcomes. *American Economic Review*, *103*(6), 2052–2086.

Heckman, J. J., & Masterov, D. V. (2007). The productivity argument for investing in young children. *Applied Economic Perspectives and Policy*, *29*(3), 446-493.

Heckman, J. J., Moon, S. H., Pinto, R. R., Savelyev, P. a., & Yavitz, A. (2010). The Rate of Return to the High/Scope Perry Preschool Program. *Journal of Public Economics*, *94*, 114–128.

Heckman, J. J., & Mosso, S. (2014). The Economics of Human Development and Social Mobility. *Annual Review of Economics*, *6*(6), 689–733.

Heckman, J. J., Pinto, R., & Savelyev, P. A. (2013). Understanding the mechanisms through which an influential early Childhood programm boosted Adult outcomes. *American Economic Review*, *103*(6), 1–35.

Hess, R. D., Holloway, S. D., Dickson, W. P., & Price, G. G. (1984). Maternal Variables as Predictors of Children' s School Readiness and Later Achievement in Vocabulary and Mathematics in Sixth Grade Author. *Child Development*, *55*(5), 1902–1912.

Ho, Andrew 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.

Ho, Andrew Dean. (2008). The Problem With "Proficiency": Limitations of Statistics and Policy Under No Child Left Behind. *Educational Researcher*, *37*(6), 351–360.

Ho, Andrew Dean. (2009). A Nonparametric Framework for Comparing Trends and Gaps Across Tests. *Journal of Educational and Behavioral Statistics*, *34*(2), 201–228.

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.

Hommel, S. L. (2013). Capacity Building for Early Childhood Development. In Britto, P. R., Engle, P. L., & Super, C. M. (Eds.), *Handbook of Early Childhood Development Research and Its Impact on Global Policy* (pp. 425–431). Oxford University Press.

Hoot, J. L., Szente, J., & Mebratu, B. (2004). Early Education in Ethiopia: Process and Prospects. *Early Childhood Education Journal, 32*(1).

Howes, C., Burchinal, M., Pianta, R., Bryant, D., Early, D., Clifford, R., & Barbarin, O. (2008). Ready to learn? Children's pre-academic achievement in pre-kindergarten programs. *Early Childhood Research Quarterly*, *23*(1), 27–50.

Hu, B. Y., Zhou, Y., Chen, L., Fan, X., & Winsler, A. (2017). Preschool expenditures and Chinese children's academic performance: The mediating effect of teacher-child interaction quality. *Early Childhood Research Quarterly*, *41*, 37–49.

Hu, L. T., & Bentler, P. M. (1999). Cutoff criteria for fit indexes in covariance structure analysis: Conventional criteria versus new alternatives. *Structural Equation Modeling*, *6*(1), 1–55.

Hungi, N., & Ngware, M. (2018). The effects of preschool participation on mathematics achievement of Grade 6 pupils in Uganda. *Educational Research for Policy and Practice*, *17*(2), 105–126.

Jakobsen, J. C., Gluud, C., Wetterslev, J., & Winkel, P. (2017). When and how should multiple imputation be used for handling missing data in randomised clinical trials - A practical guide with flowcharts. *BMC Medical Research Methodology*, *17*(1), 1–10.

Jenkins, J. M., Duncan, G. J., Auger, A., Bitler, M., Domina, T., & Burchinal, M. (2018). Boosting school readiness: Should preschool teachers target skills or the whole child? *Economics of*

Education Review, *65*, 107–125.

Johnson, R. C., & Jackson, C. K. (2018). *Reducing Inequality Through Dynamic Complementarity: Evidence from Head Start and Public School Spending*. NBER Working Papers 23489, National Bureau of Economic Research, Inc, Cambridge, MA.

García, L., Heckman, J. J., & Ziff, A. L. (2018). Gender differences in the benefits of an influential early childhood program. *European economic review, 109*, 9-22.

Jung, H., & Hasan, A. (2014). *The Impact of Early Childhood Education on Early Achievement Gaps: Evidence from the Indonesia Early Childhood Education and Development (ECED) Project.* Policy Research Working Paper 6794, Washington, D.C.: World Bank.

Seymour, P. H., Aro, M., & Erskine, J. M. (2003). Foundation literacy acquisition in European orthographies in collaboration with COST Action A8 network. *British Journal of Psychology, 93,* 143–174.

Kagan, S., Araujo, M. C., Jaimovich, A., & Aguayo, Y. C. (2015). Understanding Systems Theory and thinking: Early Childhood Education in Latin America and the Caribbean. In A. Farrell, S. L. Kagan, & K. M. Tisdall (Eds.), *The SAGE Handbook of Early Childhood Research*, 163–184.

Kagan, S. L., & Kauerz, K. (Eds.). (2012). *Early childhood systems: Transforming early learning.* New York: Teachers College Press.

Kam, C., & Frazese, R. (2005). Modeling and Interpreting Interactive Hypotheses in Regression Analysis. University of Michigan Press.

Kelly, S., & Graham, J. (2017). *The Case for Investment in Early Grade Reading*. Washington, D.C: World Bank.

Khan, Q. M., Faguet, J.-P., Gaukler, C., & Mekasha, W. (2014). Improving Basic Services for the Bottom Forty Percent: Lessons from Ethiopia. In *Improving Basic Services for the Bottom Forty Percent: Lessons from Ethiopia*. Washington, D.C: World Bank.

Kim, Y.-S. G., Boyle, H. N., Zulkowski, S. S., & Nakamura, P. (2016). *Landscape report on early grade literacy*. Washington, DC: USAID.

King, G., & Nielsen, R. (2018). Why Propensity Scores Should Not Be Used for Matching. *Political Analysis*, (617), 1–20.

Kleyn, J., Arashi, M., Bekker, A., & Millard, S. (2017). Preliminary testing of the Cobb–Douglas production function and related inferential issues. *Communications in Statistics: Simulation and Computation*, *46*(1), 469–488.

Kline, R. B. (2011). *Principles and practice of structural equation modeling, 3rd ed.* Guilford publications.

Knudsen, E. I., Heckman, J. J., Cameron, J. L., & Shonkoff, J. P. (2006). Economic, neurobiological, and behavioral perspectives on building America's future workforce. *Proceedings of the National Academy of Sciences*, *103*(27), 10155-10162.

Krafft, C. (2015). Increasing educational attainment in Egypt: The impact of early childhood care and education. *Economics of Education Review*, *46*, 127–143.

Krishnaratne, S., White, H., & Carpenter, E. (2013). Quality education for all children? What works in education in developing countries. New Delhi: International Initiative for Impact Evaluation (3ie), Working Paper, 20, 155.

Ladd, H. F., Muschkin, C. G., & Dodge, K. A. (2014). From Birth to School: Early Childhood Initiatives and Third-Grade Outcomes in North Carolina. *Journal of Policy Analysis and Management*, *33*(1), 162–187.

Layzer, J. I., & Goodson, B. D. (2006). The “quality” of early care and education settings: Definitional and measurement issues. *Evaluation review, 30*(5), 556-576.

Leak, J., Duncan, G. J., Li, W., Magnuson, K., Schindler, H., & Yoshikawa, H. (2010). Is timing everything? How early childhood education program impacts vary by starting age, program duration and time since the end of the program. In *Biennial Meeting for the Society for Research on Child Development, Montreal, Quebec, Canada.*

Lee, V. E., & Loeb, S. (1995). Where Do Head Start Attendees End Up? One Reason Why Preschool Effects Fade Out. *Educational Evaluation and Policy Analysis*, *17*(1), 62–82.

Leibowitz, A. (1974). Home Investments in Children. *Journal of Political Economy*, *82*(2), S111–S131.

Leon, J., & Singh, A. (2017). *Equating Test Scores for Receptive Vocabulary Across Rounds and Cohorts in Ethiopia, India and Vietnam*. Technical Note 40, Oxford: Young Lives.

Lerner, R.M. (1986). *Concept and theories of human development (2nd ed.)*. New York: Random House.

Lerner, R. M. (1998). Theories of human development: Contemporary perspectives. In *Handbook of child psychology.* Vol 1. New York: Wiley.

Leuven, E., & Sianesi, B. 2003. "PSMATCH2: Stata module to perform full Mahalanobis and propensity score matching, common support graphing, and covariate imbalance testing," Statistical Software Components S432001, Boston College Department of Economics.

Lewin, K. M. (2009). Access to education in sub-Saharan Africa: Patterns, problems and possibilities. *Comparative Education*, *45*(2), 151–174.

Li, M. (2012). Using the Propensity Score Method to Estimate Causal Effects: A Review and Practical Guide. *Organizational Research Methods*, 1–39.

Little, R. J. A. (1992). Regression with Missing X’s: A Review. *Journal of the American Statistical Association*, *87*(420), 1227–1237.

Little, R. J., & Rubin, D. B. (2019). *Statistical analysis with missing data* (Vol. 793). John Wiley & Sons.

Little, R. J., & Rubin, D. B. (2002). Maximum likelihood for general patterns of missing data: Introduction and theory with ignorable nonresponse. *Statistical analysis with missing data*, 164-189.

Livingston, S. A. (2006). Double P-P Plots for Comparing Differences between Two Groups. *Journal of Educational and Behavioral Statistics*, *31*(4), 431–435.

Loeb, S., Bridges, M., Bassok, D., Fuller, B., & Rumberger, R. W. (2007). How much is too much? The influence of preschool centers on children’s social and cognitive development. *Economics of Education Review*, *26*(1), 52–66.

Long, J. S., & Freese, J. (2014). *Regression models for categorical dependent variables using Stata*. Stata press.

Lopez Boo, F. (2016). Socio-economic status and early childhood cognitive skills: A mediation analysis using the Young Lives panel. *International Journal of Behavioral Development*, *40*(6), 500–508.

Lorenz, M. O. (1905). *Methods of Measuring the Concentration of Wealth.* Publications of the American Statistical Association, 9, 209-219.

Ludwid, J., & Miller, D. (2007). Does Head Start Improve Children’s Life Chances? Evidence From A Regrerssion Discontinuity Design. *The Quarterly journal of economics, 122(1), 159-208.*

MacBeath, J. (2010). Self-evaluation for school improvement. In *Second international handbook of educational change* (pp. 901-912). Springer, Dordrecht.

Magnuson, K. A., & Duncan, G. J. (2017). Can Early Childhood Interventions Decrease Inequality of Economic Opportunity? *RSF: The Russell Sage Foundation Journal of the Social Sciences*, *2*(2), 123.

Magnuson, K. A., Kelchen, R., Duncan, G. J., Schindler, H. S., Shager, H., & Yoshikawa, H. (2016). Do the effects of early childhood education programs differ by gender? A meta-analysis. *Early Childhood Research Quarterly*, *36*, 521–536.

Magnuson, K. A., Meyers, M. K., & Ruhm, C. J. (2004). Inequality in Preschool Education and School Readiness. *American Educational Research Journal*, *41*(1), 115–157.

Magnuson, K. A., Ruhm, C., & Waldfogel, J. (2007). The persistence of preschool effects: Do subsequent classroom experiences matter? *Early Childhood Research Quarterly*, *22*(1), 18–38.

Mantel, N., & Haenszel, W. (1959). Statistical aspects of the analysis of data from retrospective studies of disease. *Journal of the National Cancer Institute, 22*, 719-48.

Marmot, M., & Bell, R. (2012). Fair society, healthy lives. *Public health, 126*, S4-S10.

Martinez, S., Naudeau, S., & Pereira, V. (2012). T*he Promise of Preschool in Africa : A Randomized Impact Evaluation of Early Childhood Development in Rural Mozambique*. ECD Report, Washignton D.C.World Bank; Save the Children.

Mashburn, A. J., Pianta, R. C., Hamre, B. K., Downer, J. T., Barbarin, O., Bryant, D., ... Howes, C. (2008). Measures of classroom quality in prekindergarten and children's development of academic, language, and social skills. *Child Development*, *79*(3), 732–749.

McCoy, D. C., Morris, P. A., Connors, M. C., Gomez, C. J., & Yoshikawa, H. (2016). Differential effectiveness of Head Start in urban and rural communities. *Journal of applied developmental psychology*, *43*, 29-42.

McCoy, D. C., Peet, E. D., Ezzati, M., Danaei, G., Black, M. M., Sudfeld, C. R., … Fink, G. (2016). Early Childhood Developmental Status in Low- and Middle-Income Countries: National, Regional, and Global Prevalence Estimates Using Predictive Modeling. *PLoS Medicine*, *13*(6), 1–

18.

McCoy, D. C., Yoshikawa, H., Ziol-Guest, K. M., Duncan, G. J., Schindler, H. S., Magnuson, K., … Shonkoff, J. P. (2017). Impacts of Early Childhood Education on Medium- and Long-Term Educational Outcomes. *Educational Researcher*, *46*(8), 474–487.

McCoy, D. C., Zuilkowski, S. S., Yoshikawa, H., & Fink, G. (2017). Early Childhood Care and Education and School Readiness in Zambia. *Journal of Research on Educational Effectiveness, 10*(3), 482–506.

McEwan, P. J. (2015). Improving learning in primary schools of developing countries: A meta-analysis of randomized experiments. *Review of Educational Research, 85*(3), 353-394.

Michalopoulos, C., Bloom, H. S., & Hill, C. J. (2004). Can Propensity-Score Methods Match the Findings from a Random Assignment Evaluation of Mandatory Welfare-to-Work Programs? *Review of Economics and Statistics, 86(1), 156-179.*

Ministry of Education. (2010a). *Education Sector Development Program IV (ESDP IV) 2010/2011 - 2014/2015.* Addis Ababa: Government of Ethiopia..

Ministry of Education. (2010b). *National Policy Framework for Early Childhood Care and Education*. Addis Ababa: Government of Ethiopia.

Ministry of Education. (2002). *The Education and Training Policy and Its Implimentation*. Addis Ababa: Government of Ethiopia.

Ministry of Education. (2010). *Education Statistics Annual Abstract 2002 E.C. (2009/10)*. Addis Ababa: Government of Ethiopia.

Ministry of Education. (2012). *Education Statistics Annual Abstract 2004 E.C. (2011/12)*. Addis Ababa: Government of Ethiopia.

Ministry of Education. (2015). *Education Sector Development Programme V (ESDP V) 2015/16 - 2019/20 G.C.* (ESDP V), 2008–2012. Addis Ababa: Government of Ethiopia.

Ministry of Education. (2017). *Education Statistics Annual Abstract 2008 E.C. (2015/16)*. Addis Ababa: Government of Ethiopia.

Moloi, M. Q., and Chetty, M. (2010). *The SACMEQ III project in South Africa: A study of the conditions of schooling and the quality of education*. Pretoria: South Africa Ministry of Basic Education.

Morgan, S. L., & Winship, C. (2015). *Counterfactuals and causal inference*. Cambridge University Press.

Murnane, R., & Willett, J. (2010). *Methods Matter: Improving Causal Inference in Educational and Social Science Researc*h. Oxford University Press.

Muschkin, C. G., Ladd, H. F., Dodge, K. A., & Bai, Y. (2018). Gender Differences in the Impact of North Carolina’s Early Care and Education Initiatives on Student Outcomes in Elementary School. *Educational Policy*, 0895904818773901.

Mwaura, P. A. M., Sylva, K., & Malmberg, L. (2008). Evaluating the Madrasa preschool programme in East Africa: a quasi-experimental study. *International Journal of Early Years Education*, *16*(3), 237–255.

Nagin, D. S., & Tremblay, R. E. (2001). Analyzing developmental trajectories of distinct but related behaviors: a group-based method. *Psychological Methods*, *6*(1), 18–34.

Nagy, W. ., & Scott, J. . (2000). Vocabulary processes. In Lawrence Erlbaum Associates Publishers. (Ed.), *Handbook of reading research, Vol. 3* (pp. 269–284).

National Educational Assessment and Examinations Agency. (2016). *Ethiopian Fifth National Learning Assessment of Grades 4 and 8 students*. Addis Ababa, NEAEA.

National Institute of Child Health and Human Development Early Child Care Research Network (NICHD), & Duncan, G. J. (2003). Modeling the Impacts of Child Care Quality on Children ' s Preschool Cognitive Development. *Child Development*, *74*(5), 1454–1475.

Naudeau, S., Kataoka, N., Valerio, A., Neuman, M., & Elder, L. (2011). *Investing in young children: An early childhood development guide for policy dialogue and project preparation.* Washington, D.C.: The World Bank.

Nores, M., & Barnett, W. S. (2010). Benefits of early childhood interventions across the world: (Under) Investing in the very young. *Economics of Education Review*, *29*(2), 271–282.

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.

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

Peng, S. S., & Wright, D. (1994). Explanation of Academic Achievement of Asian American Students. *The Journal of Educational Research*, *87*(6), 346–352.

Pepe, M. S. 2003. *The Statistical Evaluation of Medical Tests for Classification and Prediction*. New York: Oxford University Press.

Pforr, K. (2013). How to interpret the logistic regression with fixed effects. Presentation prepared for the *5th ESRA Conference*.

Phillips, D., Lipsey, M., Dodge, K. A., Haskins, R., Bassok, D., Burchinal, M. R., … Weiland, C. (2017). *Puzzling It Out: The Current State of Scientific Knowledge on Pre-Kindergarten Effects:* A Consensus Statement. Issues in pre-kindergarten programs and policy, 19-30.

Pianta, Howes, Burchinal, M., Bryant, D., Clifford, R., & Early, D. (2005). Features of Pre-Kindergarten Programs, Classrooms, and Teachers: Do They Predict Observed Classroom Quality and Child–Teacher Interactions? *Applied Developmental Science, 9*(3), 144–159.

Piper, B. (2010). *Ethiopia Early Grade Reading Assessment. Data Analytic Report: Language and Early Learning Ethiopia*. Addis Ababa: USAID/Ethiopia.

Piper, B., Simmons Zuilkowski, S., Dubeck, M., Jepkemei, E., & King, S. J. (2018). Identifying the essential ingredients to literacy and numeracy improvement: Teacher professional development and coaching, student textbooks, and structured teachers' guides. *World Development, 106,* 324–336.

Piper, B., Sitabkhan, Y., Mejia, J., & Betts, K. (2018). *Effectiveness of Teachers' Guides in the Global South: Scripting, Learning Outcomes, and Classroom Utilization*. RTI Occasional Paper, Research Triangle Park, NC: RTI Press.

Piper, B., Zuilkowski, S. S., & Mugenda, A. (2014). Improving reading outcomes in Kenya: First-year effects of the PRIMR Initiative. *International Journal of Educational Development*, *37*, 11–21.

Preacher, K. J., Zyphur, M. J., & Zhang, Z. (2010). A general multilevel SEM framework for assessing multilevel mediation. *Psychological Methods*, *15*(3), 209–233.

Puma, M., Bell, S., Cook, R., Heid, C., Shapiro, G., Broene, P., ... & Ciarico, J. (2010). Head Start Impact Study. Final Report. *US Administration for Children & Families*.

Ramey, C. T., & Ramey, S. (1998). Early Intervention and Early Experience. *American Psychologist*, *53*(2), 109–120.

Rao, N., Sun, J., Pearson, V., Pearson, E., Liu, H., Constas, M. A., & Engle, P. L. (2012). Is Something Better Than Nothing? An Evaluation of Early Childhood Programs in Cambodia. *Child Development*, *83*(3), 864–876.

Rao, N., Sun, J., Wong, J. M. S., Weekes, B., Ip, P., Shaeffer, S., … Lee, D. (2014). *Early childhood development and cognitive development in developing countries: A rigorous literature review*. Department for International Development. U.K.

Rao, N., Sun, J., Zhou, J., & Zhang, L. (2012). Early achievement in rural China: The role of preschool experience. *Early Childhood Research Quarterly*, *27*(1), 66–76.

Raudenbush, S. W., & Bryk, A. S. (2002). *Hierarchical linear models : applications and data analysis methods* (2nd ed.). Thousand Oaks, CA: Sage Publications, Inc.

Reardon, S. F., & Portilla, X. A. (2016). Recent trends in income, racial, and ethnic school readiness gaps at kindergarten entry. *Aera Open*, *2*(3), 2332858416657343.

Reynolds, A. J., Ou, S., & Topitzes, J. W. (2004). Paths of Effects of Early Childhood Intervention on Educational Attainment and Delinquency: A Confirmatory Analysis of the Chicago Child-Parent Centers. *Child Development*, *75*(5), 1299–1328.

Reynolds, C. L., & Desjardins, S. L. (2009). The Use of Matching Methods in Higher Education Research: Answering Whether Attendance at a Two-Year Institution Results in Differences in Educational Attainment. *Higher Education: Handbook of Theory and Research*, *24*, 47–97.

Richter, L. M., Daelmans, B., Lombardi, J., Heymann, J., Boo, F. L., Behrman, J. R., … Darmstadt, G. L. (2017). Investing in the foundation of sustainable development: pathways to scale up for early childhood development. *The Lancet*, *389*(10064), 103–118.

Rolleston, C., James, Z., & Aurino, E. (2013). *Exploring the effect of educational opportunity and inequality on learning outcomes in Ethiopia, Peru, India, and Vietnam*. Background paper prepared for the Education for All Global Monitoring Report 2013/4. Paris: UNESCO.

Rosenbaum, P. R. (2005). Sensitivity Analysis in Observational Studies. In B. S. Everitt & D. C. Howell (Eds.), *Encyclopedia of Statistics in Behavioral Science* (Vol. 4, pp. 1809–1814). Chichester: John Wiley & Sons, Ltd.

Rosenbaum, P. R., & Rubin, D. B. (1983). The central role of the propensity score in observational studies for causal effects. *Biometrika*, *70*(1), 41–55.

Rosenbaum, P. R., & Rubin, D. B. (1985). Constructing a Control Group Using Multivariate Matched Sampling Methods That Incorporate the Propensity Score. *The American Statistician, 39*(1), 33.

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. http://dx.doi.org/10.4135/9781526402004.n14

RTI. (2015). *Results of the Early Grade Reading Benchmarking Workshop in Ethiopia*. Addis Ababa: Commissioned by USAID/Ethiopia.

Rubin, D B. (1997). Estimating causal effects from large data sets using propensity scores. *Annals of Internal Medicine*, *127*(82), 757–763.

Rubin, Donald B. (1976). Inference and Missing Data. *Biometrika*, *63*(3), 581.

Rubin, Donald B. (1996). Multiple Imputation After 18+ Years. *Journal of the American Statistical Association*, *91*(434), 473.

Ruhm, C. R., & Waldfogel, J. (2012). Long-Term Effects of Early Childhood Care and Education. *Nordic Economic Policy Review, 1*(6149), 23–51.

Rutter, M. (2006). *Genes and behavior: Nature-nurture interplay explained.* Blackwell Publishing.

Sabatier, P. A., & Jenkins-Smith, H. C. (1993). *Policy change and learning: An advocacy coalition approach*. Westview Pr.

Sameroff, A. J. (1983). Developmental systems: Contexts and evolution. In Kessen. W. (ed.) *Handbook of child psychology: Vol. 1. History, theory, and methods*. pp. 237 – 294. New York: Wiley.

Sammons, P. (2009). The dynamics of educational effectiveness: a contribution to policy, practice and theory in contemporary schools. *School Effectiveness and School Improvement*, *20*(1), 123–129.

Sandefur, G. D., Meier, A. M., & Campbell, M. E. (2006). Family resources, social capital, and college attendance. *Social Science Research*, *35*(2), 525–553.

Sayre, R. K., Devercelli, A. E., Neuman, M. J., & Wodon, Q. (2015). *Investing in Early Childhood Development: Review of the World Bank's Recent Experience*. Washingtong D.C.: The World Bank.

Schady, N., Behrman, J., Caridad Araujo, M., Azuero, R., Bernal, R., Bravo, D., … Vakis, R. (2015). Wealth Gradients in Early Childhood Cognitive Development in Five Latin American Countries. *Journal of Human Resources*, *50*(2), 446–463.

Schultz, T. (1961). Investment in human capital. *The American Economic Review*, 1-17.

Schweinhart, L., Montie, J., Xiang, Z., Barnett, W. S., Belfield, C. R., & Nores, M. (2005). Lifetime effects: The High/Scope Perry Preschool Study through age 40. *Childhood Education, 8*3(2), 121.

Seid, Y. (2016). Does learning in mother tongue matter? Evidence from a natural experiment in Ethiopia. *Economics of Education Review*, *55*, 21–38.

Sylva, K., Siraj-Blatchford, I., Taggart, B., Sammons, P., Melhuish, E., Elliot, K., & Totsika, V. (2006). Capturing quality in early childhood through environmental rating scales. *Early childhood research quarterly*, *21*(1), 76-92.

Phillips, D. A., & Shonkoff, J. P. (Eds.). (2000). *From neurons to neighborhoods: The science of early childhood development*. Washington D.C.: National Academies Press.

Silva, J. M. C. S., & Kemp, G. (2016). Partial effects in fixed effects models. In *22nd London Stata Users Group Meeting* (pp. 1–18).

Singh, A. (2015, April). Learning more with every year: School year productivity and international learning divergence. In *CESifo Area Conference on the Economics of Education, CESifo Group, Munich, September* (pp. 11-12).

Singh, R., & Mukherjee, P. (2018). Effect of preschool education on cognitive achievement and subjective wellbeing at age 12: evidence from India. *Compare*, *7925*, 1–19.

Smith, J. A., & Todd, P. E. (2005). Does matching overcome LaLonde's critique of nonexperimental estimators?. *Journal of Econometrics*, *125*(1-2), 305-353.

Snijders, T. A. B., & Bosker, R. J. (1999). *Multilevel analysis: an introduction to basic and advanced multilevel modeling*. Sage, London.

Snilstveit, B., Stevenson, J., Menon, R., Phillips, D., Gallagher, E., Geleen, M., … Jimenez, E. (2016). *The impact of education programmes on learning and school participation in low- and middle-income countries: a systematic review summary report*, 3ie Systematic Review Summary 7. London: International Initiative for Impact Evaluation (3ie).

Sobel, M. E. (1982). Asymptotic Confidence Intervals for Indirect Effects in Structural Equation Models. *Sociological Methodology*, *13*(1982), 290–312.

Spencer, B. D. . (1983). On Interpreting Test Scores as Social Indicators : Statistical Considerations. *Journal of Educational Measurement*, *20*(4), 317–333.

Stanovich, K. E. (1986). Matthew Effects in Reading: Some Consequences of Individual Differences in the Acquisition of Literacy. *Reading Research Quarterly*, *21*(4), 360–407.

Stock, J. H., & Watson, M. W. (2015). *Introduction to Econometrics*. Pearson.

Sun, J., Rao, N., & Pearson, E. (2015). *Policies and Strategies to Enhance the Quality of Early Childhood Educators*. Background paper prepared for the Education for All Global Monitoring Report 2015. Paris: UNESCO.

Teferra, T. & Hagos, B. (2016). *Assessment of the Status of O-Class in Four Regional States of Ethiopia*. Commissioned by World Bank, Addis Ababa: Institute of Education, Health and Development.

Tesema, M. T., & Braeken, J. (2018). Regional inequalities and gender differences in academic achievement as a function of educational opportunities: Evidence from Ethiopia. *International Journal of Educational Development*, *60*(August 2017), 51–59.

Tesfay, N., & Malmberg, L.-E. (2014). Horizontal inequalities in children's educational outcomes in Ethiopia. *International Journal of Educational Development*, *39*, 110–120.

Thelen, E., & Smith, L. B. (1994). *A dynamic systems approach to the development of cognition and action*. Cambridge, MA: Bradford Books/MIT Press.

Thomason, A. C., & La Paro, K. M. (2009). Measuring the quality of teacher–child interactions in toddler child care. *Early Education and Development, 20*(2), 285-304.

Todd, P. E., & Wolpin, K. I. (2003). On the Specification and Estimation of the Production Function for Cognitive Achievement. *The Economic Journal*, *113*(485), F3–F33.

Todd, P. E., & Wolpin, K. I. (2007). The Production of Cognitive Achievement in Children: Home, School, and Racial Test Score Gaps. *Journal of Human Capital*, *1*(1), 91–136.

UNDP. (2018). Human Development Indices and Indicators. 2018 Statistical Update. *United Nations Development Programme*, *27*(4), 123.

UNESCO. (2000). *The Dakar Framework for Action, Education for All: Meeting our Collective Commitments* (pp. 26–28). Paris: UNESCO.

UNESCO. (2007). EFA Global Monitoring Report 2007: Strong Foundations. In *Early Childhood Care & Education*. Paris: UNESCO.

UNESCO. (2014). EFA Global Monitoring Report 2013/4: Teaching and Learning. In *Achieving quality for all*. Paris: UNESCO.

UNESCO. (2015). *Thematic Indicators to Monitor the Education 2030 Agenda*. Paris: UNESCO.

UNICEF. (2016). *Journeys to Scale: Accompanying the Finalists of the Innovations in Education Initiative.* New York: UNICEF.

Van der 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*. Final Report for the Department of Basic Education and the Department of Performance Monitoring and Evaluation in the Presidency. University of Stellenbosch, Stellenbosch.

Vandemoortele, M. (2018). *Inequality in attainment from early childhood to adolescence longitudinal evidence from Ethiopia.*Working Paper 177, Oxford: Young Lives.

Vargas-Barón, E. (2015). *Policies on early childhood care and education: their evolution and some impacts.* Paris: UNESCO.

Von Bertalanffy, L. (1968). *General system theory: Foundations, development, applications.* New York: George Braziller.

Walberg, H. J., & Tsai, S.-L. (1983). Matthew Effects in Education. *American Educational Research Journal*, *20*(3), 359–373.

Walker, S. P., Chang, S. M., Vera-Hernandez, M., & Grantham-McGregor, S. (2011). Early Childhood Stimulation Benefits Adult Competence and Reduces Violent Behavior. *Pediatrics*, *127*(5), 849–857.

Watamura, S. E., Morrissey, T. W., Phillips, D. A., Mccartney, K., & Bub, K. (2011). Double Jeopardy : Poorer Social-Emotional Outcomes for Children in the NICHD SECCYD Experiencing Home and Child-Care Environments That Confer Risk. *Child Development*, *82*(1), 48–65.

Weiland, C., & Yoshikawa, H. (2013). Impacts of a prekindergarten program on children's mathematics, language, literacy, executive function, and emotional skills. *Child Development*, *84*(6), 2112–2130.

Wen, Z., & Fan, X. (2015). Monotonicity of effect sizes: Questioning kappa-squared as mediation effect size measure. *Psychological Methods*, *20*(2), 193–203.

Wilk, A. M. B., & Gnanadesikan, R. (1968). Probability Plotting Methods for the Analysis of Data. *Biometrika*, *55*(1), 1–17.

Williams, R. (2012). Using the margins command to estimate and interpret adjusted predictions and marginal effects. *The Stata Journal*, *12*(2), 308–331.

Williams, R. (2018). *Panel Data and Multilevel Models for Categorical Outcomes: Fixed effects and conditional logit models*. 1–5. Retrieved from https://www3.nd.edu/~rwilliam/

Wilson, I., Huttly, S. R. A., Fenn, B., Wilson, I., Huttly, S. R. A., Fenn, B., … Fenn, B. (2006). A Case Study of Sample Design for Longitudinal Research : Young Lives. *International Journal of Social Research Methodology, 9*(5), 351–365.

Woldehanna, T. (2011). The Effects of Early Childhood Education Attendance on Cognitive Development: Evidence from Urban Ethiopia. *Ethiopian Journal of Economics*, *20*(1), 113–164.

Woldehanna, T. (2016). Inequality, preschool education and cognitive development in Ethiopia. *International Journal of Behavioral Development*, *40*(6), 509–516.

Woldehanna, T., Amha, W., Yonis, M. B., & Tafere, Y. (2018). *Understanding and supporting sustained pathways out of extreme poverty and deprivation in Ethiopia: National Synthesis Report*. Addis Ababa. Association of Ethiopian Micro-Finance Institutions.

Woldehanna, T., & Araya, M. W. (2017). *Early Investment in Preschool and Completion of Secondary Education in Ethiopia: Lessons From Young Lives.* Working Paper 168, Oxford: Young Lives.

Woldehanna, T., & Gebremedhin, L. (2012). *The Effects of Pre-school Attendance on the Cognitive Development of Urban Children aged 5 and 8 Years: Evidence from Ethiopia*. Working Paper 89, Oxford: Young Lives.

Woldehanna, T., Jones, N., & Tefere, B. (2005). C*hildren's Educational Completion Rates and Achievement: Implications for Ethiopia's Second Poverty Reduction Strategy (2006–10)*. London: Young Lives/Save the Children UK.

Wolf, S., Aber, L. J., & Morris, P. A. (2013). Drawing on psychological theory to understand and improve antipoverty policies: The case of conditional cash transfers. *Psychology, Public Policy, and Law*, *19*(1), 3–14.

Wong, V., Cook, T., Barnett, W. S., & Jung, K. (2008). An Effectiveness-Based Evaluation of Five State Pre-Kindergarten Programs. *Journal of Policy Analysis and Management*, *27*(1), 122–154.

Woodhead, M. (2009). *Pathways through early childhood education in Ethiopia, India and Peru: Rights, equity and diversity*. Young Lives Working Paper No. 54. Oxford: Young Lives.

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

Bernard van Leer Foundation.

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.

World Bank. (2016). *Striving for Excellence: Analysis of Ethiopia National Learning Assessments 2011 and 2015*. Unpublished report. Washington, D.C.: World Bank.

World Bank. (2017). *Program Appraisal Document, General Education Quality Improvement Program for Equity (GEQIP-E)*. Washington, D.C.: World Bank.

Wuermli, A., Silbereisen, R. K., Lundberg, M., Lamont, M., Behrman, J. R., & Aber, L. (2012). A Conceptual Framework. In M. Lundberg & A. Wuermli (Eds.), *Children and Youth in Crisis* (pp. 29-102).

Yilmaz, S., & Venugopal, V. (2008). *Local government discretion and accountability in Ethiopia* (No. paper0838). International Center for Public Policy, Andrew Young School of Policy Studies, Georgia State University.

Yoshikawa, H., & Hsueh, J. (2001). Child development and public policy: Toward a dynamic systems perspective. *Child Development*, *72*(6), 1887–1903.

Yoshikawa, H., & Nieto, A. M. (2013). Paradigm Shifts and New Directions in Research on Early Childhood Development Programs in Low- and Middle-Income Countries. In Britto, P. R., Engle, P. L., & Super, C. M. (Eds.), *Handbook of Early Childhood Development Research and Its Impact on Global Policy* (pp. 487–500).

Yoshikawa, H., Weiland, C., Brooks-gunn, J., Burchinal, M. R., Espinosa, L. M., Gormley, W. T., … Zaslow, M. J. (2013). *Investing in our future: The evidence base on preschool education*. Society for Research in Child Development and the Foundation for Child Development.

Yoshikawa, H., Wuermli, A. J., Raikes, A., Kim, S., & Kabay, S. B. (2018). Toward High-Quality Early Childhood Development Programs and Policies at National Scale: Directions for Research in Global Contexts. *Social Policy Report*, *31*(1), 1–36.

Young Lives. (2018). *Young Lives Survey Design and Sampling (Round 5): Ethiopia*. Retrieved from https://www.younglives.org.uk/sites/www.younglives.org.uk/files/ETHIOPIA-SurveyDesign-Factsheet-Jan18_0.pdf

Zhai, F., Brooks-Gunn, J., & Waldfogel, J. (2014). Head Start’s impact is contingent on alternative type of care in comparison group. *Developmental psychology*, *50*(12), 2572.

APPENDICES

Appendix A. [Chapter 2] Robustness Check: Propensity Score Matching

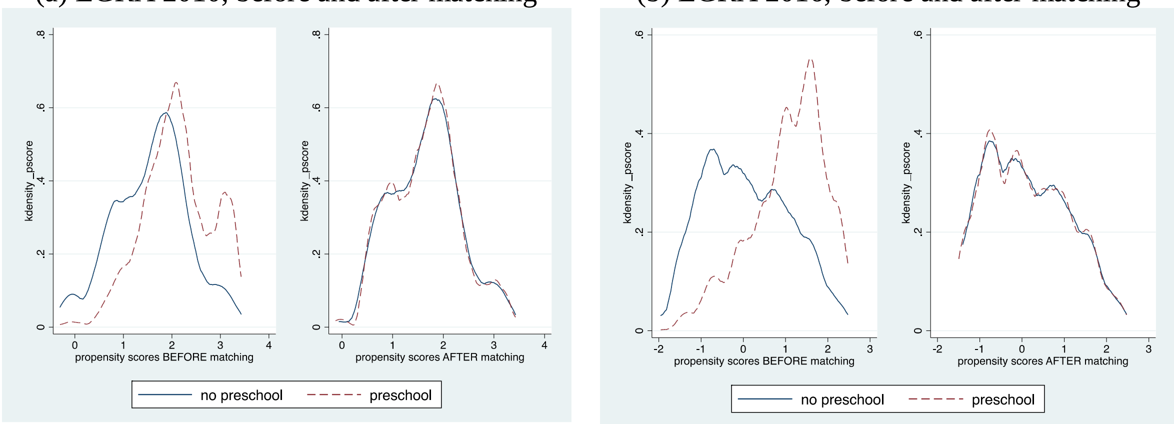
As presented in Table 2.12 above, I used the propensity score matching as a robustness check for the results derived from the OLS and school fixed effects model. The PSM, developed by Rubin and Rosenbaum (1983), is a widely used quasi-experimental method in observational studies when conducting an experiment is not feasible. Specifically, PSM emulates a situation which experimental research achieves through randomization by modeling the treatment assignment patterns directly and creating sub-groups which match in their likelihood of belonging to either a treatment or a control group (technical details appear in Chapter 4; Guo & Fraser, 2015). Here, the propensity score is defined as the conditional probability of attending preschool (treatment) given pre-treatment characteristics of preschool attendees and non-attendees. By accounting for observable sources of non-random selection into preschool, this approach allowed me to identify an adequate counterfactual, which compared children who attended preschool (treated group) with those who shared similar socio-demographic characteristics but did not attend preschool (control group).

Specifically, I used a kernel density matching approach (Heckman, Ichimura, & Todd, 1998) to match preschool attendees and non-attendees, drawing from the same set of covariates used in the OLS and school fixed effects model. Kernel matching is a nonparametric matching approach that compares the outcome of each individual who attended preschool to a weighted average of the outcomes of all children who didn’t attend preschool, with the highest weight being placed on those with propensity scores nearest to the particular preschool attendee. Kernel matching uses more information for each match, thus producing a lower variance than traditional propensity scores matching techniques.

As a first step to create a reliable counterfactual, Figure 2.A presents the common support areas (i.e., probability densities), especially before and after PSM. This figure shows the extent to which treated (preschool) and non-treated (non-preschool) students have comparable observed likelihoods of attending preschool. Notably, before kernel-based matching (left panel of each cohort), the overlapped areas in the distribution of the propensity score declines from 2010 to 2016, indicating that the gaps widen between preschool attendees and non-attendees. After kernel-based matching (right panel of each cohort), the projection presents a great deal of overlap between the two groups, which supports establishing the comparability of the treated and untreated group.

Figure A. 1. Common Support Area for Preschool Attendees and Non-Attendees

(a) EGRA 2010, before and after matching (b) EGRA 2016, before and after matching

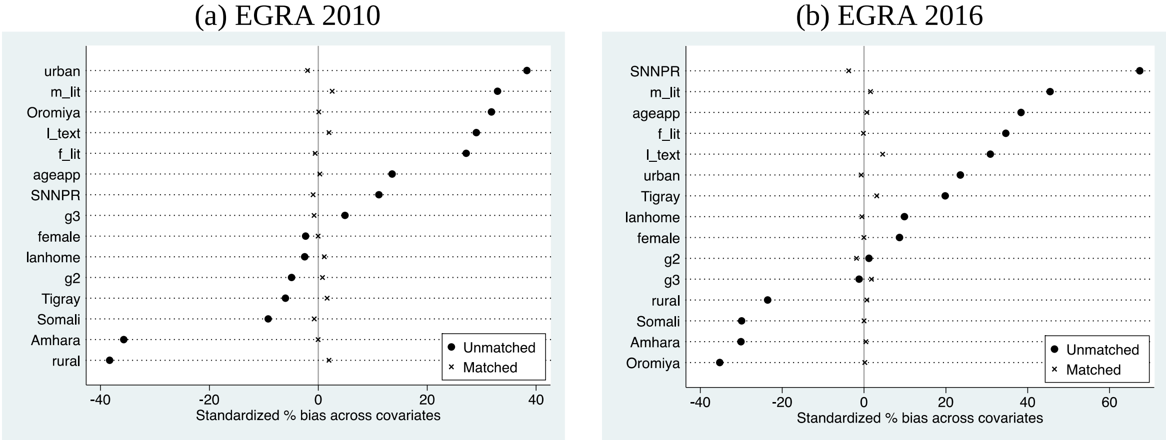


Note: In EGRA 2010, 61 students are out of common support (7,876 no-pre; 1,184 pre; 9,060 total); In EGRA 2016, 114 students are out of common support (5,316 no-pre; 2,862 pre; 8,178 total).
Source: EGRA Dataset 2010, 2016, USAID

Second, to meet the covariate balance assumption of PSM, I checked the balance in observed characteristics between preschool attendees and non-attendees through the matching process. Figure 2.B is a visual presentation of standardised differences and associated percentage bias by unmatched and matched groups (Caliendo & Kopeinig, 2008). This figure captures how matching procedure contributes to the convergence of associated percentage bias into zero in each of covariates. According to the threshold set to 0.20 (Rosenbaum & Rubin, 1985), standardised percentage bias across covariates (observed characteristics) displays excellent balance, closer to zero, for both EGRA cohorts.

Figure A. 2. Covariate Balance between Matched and Unmatched Groups

(a) EGRA 2010 (b) EGRA 2016



Source: EGRA Dataset 2010, 2016, USAID

Finally, Table 2.A shows the results of PSM with each value of treated and control groups by region. As described earlier, most of the results are similar or slightly larger than estimates of OLS and school fixed effect model. This corroborates the finding of the present study that the role of preschool has been strengthened during a large-scale expansion of public preschool in predicting students' early grade reading achievement.

Table A. 1. Results of Propensity Score Matching

ORF (*correct words per minute*)	2010 EGRA			2016 EGRA		
	Treated	Control	Diff (t-test)	Treated	Control	Diff (t-test)
Total Average	22.87	20.89	1.98**	22.03	18.94	3.09***
Tigray	25.60	20.66	4.94***	19.84	19.49	0.35
Amhara	32.02	27.20	4.82**	37.27	32.70	4.57***
Oromia	24.41	22.30	2.11	19.13	15.04	4.09***
Somali	32.20	33.97	-1.77	14.60	9.80	4.80***
SNNP(s)	8.74	8.72	0.02	20.62	19.06	1.56*

Note: All estimates include sampling weight. *** $p<0.01$, ** $p<0.05$, *$p<0.1$
Source: EGRA Dataset 2010, 2016, USAID

Appendix B. [Chapter 3] Robustness Check: Interaction Effects of Woreda Poverty Level

In this sub-section, I turn to testing the robustness of my estimates to alternative explanations. I reported above that the benefits of preschool were particularly pronounced for children from advantaged backgrounds—those living in urban areas and/or who have a literate father—than their peers from disadvantaged backgrounds. The dichotomy variables such as urbanicity and parental literacy provide useful information for identifying a particular group for whom the benefits of preschool were greater; however, it has limited ability to capture a more detailed picture of the socioeconomic gradients that may affect the preschool benefits. To address this limitation, I used the 'woreda poverty index' drawn from the Ethiopia Socio-economic Survey 2015/16 (ESS) (Central Statistical Agency & World Bank, 2017) for my robustness check.[167] A woreda, which is equivalent to a district, is the third-level sub-national administrative division in Ethiopia, following region and zone. There currently are approximately 670 rural woreda and 100 urban woreda in Ethiopia (Yilmaz & Venugopal, 2008).

ESS 2015/16 is a nationally representative survey of more than 3,600 households in Ethiopia in both rural and urban areas. The objective of ESS is to collect multi-topic, household-level panel data, with a special focus on improving agriculture statistics and generating a clearer understanding of its connection with welfare indicators and socioeconomic characteristics (CSA & World Bank, 2017). Drawing on the ESS 2015/16, the national-level poverty index was initially estimated based on the 'per adult total consumption expenditure' of each household, which consists of food consumption and non-food consumption expenditures (Woldehanna, Amha, Yonis, & Tafere, 2018).[168] Per-adult total consumption, for example, captured the stark urban-rural difference, as household consumption levels in urban areas were more than 2.5 times higher than those in rural areas. In the present study, the woreda-level poverty index was estimated by following the same procedure used for the national figure. However, in spite of the advantage of using a detailed poverty indicator, there are two main constraints that require extra caution in interpreting the results of this analysis. First, ESS is only representative at the national level, not at the regional or woreda levels. There is a

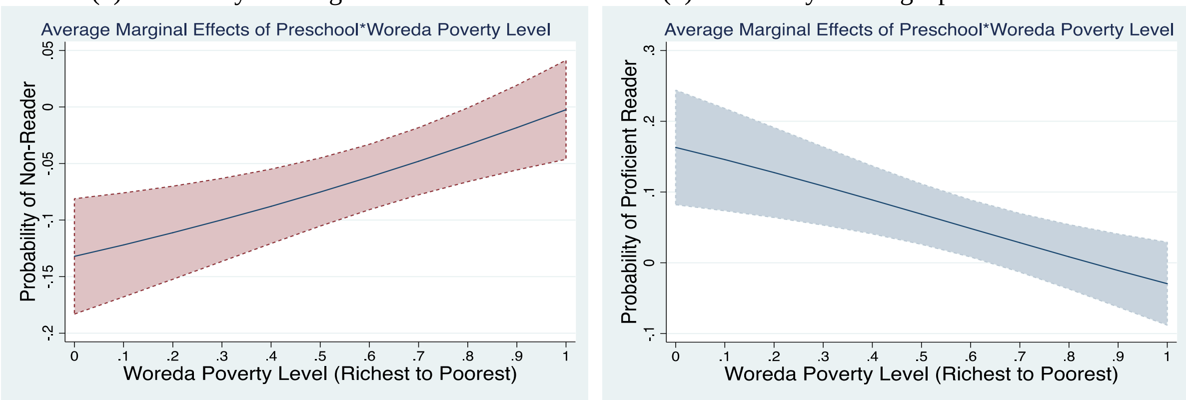
[167] The ESS is a collaborative project between the Central Statistics Agency of Ethiopia and the World Bank Living Standards Measurement Study-Integrated Surveys of Agriculture.
[168] The food consumption report was based on a recall period of seven days, then multiplied by 52 to convert it into a yearly consumption value.

possibility that, in some woredas, rich households were oversampled, and vice versa. While the woreda poverty index is highly correlated with the urban and rural location ($d = 0.37$, $p < 0.001$), any generalisation of the results should be avoided. Second, when I matched the EGRA and ESS datasets, only 80 percent (179 out of 225) of EGRA 2016 schools are located in the woreda where the ESS 2015/16 survey was conducted, thus there was a partial loss of the EGRA sample. Considering these two major limitations, I used this indicator only for the robustness check on the differential influence of preschool by urban and rural areas.

Woreda poverty levels, based on the per-adult total consumption expenditure of households in the woreda, starts from '0' for the richest to '1' for the poorest population. Figure A-1 presents the average marginal effects of the association between preschool and student outcomes on (1) the probability of being a non-reader, and (2) the probability of being a proficient reader, as a function of the woreda poverty-level index. Consistent with the differential influence by urban and rural location, the results showed that the benefits of preschool were much greater for children from affluent backgrounds than poorer backgrounds. For children living in the richest woreda, attending a preschool reduced the probability of being a non-reader by 13.2 percentage points ($p < 0.01$), whereas there were virtually no measurable benefits for children living in the poorest woreda ($d = -0.02$, $p > 0.01$). Similarly, preschool had a more discernible benefit for children from the richest woreda on the probability of being a proficient reader, which increased by 16.3 percentage points ($p < 0.01$), but there were no significant gains from preschool for children living in the poorest woreda ($d = -0.01$, $p > 0.01$). The results of this analysis reaffirm that the large-scale expansion of preschool in Ethiopia may not play an 'equaliser' role for learning inequality, as intended. Apparently, the benefits of preschool were concentrated on children from wealthier backgrounds rather than on their peers from poorer and marginalised backgrounds.

Figure B. 1. EGRA 2016: Interaction between Preschool and Woreda Poverty Level

(1) Probability of being a non-reader (2) Probability of being a proficient reader



Note: The areas along with the line represent the upper and lower bounds of the confidence interval. Confidence level = 0.95.
Source: EGRA Dataset 2010, 2016, USAID

Appendix C. [Chapter 4] Propensity Score Matching Sensitivity Checks to Unobserved Variables

To assess the extent to which the current study is sensitive to hidden selection bias, I present the results of a Rosenbaum bounds sensitivity analysis (Rosenbaum, 2005) and Mantel-Haenszel bounds sensitivity analysis (Becker & Caliendo, 2007). In testing a model's sensitivity, the basic question is whether unobserved factors can alter the relationship between treatment and outcomes. These two sensitivity tests determine how strongly confounding factors—that is, unobserved variables that might influence both assignment to preschool and students' later educational outcomes—would need to influence the selection process of forming relationships to undermine the implications of the matching analysis. The bounding approach does not directly test the assumption that there is no confounding variable in the model but provides evidence on 'the degree to which any significance results hinge on this untestable assumption' (Bharath et al., 2011, p.1197).

For the models with continuous outcome variables (e.g., test scores), Stata's *rbounds* programme (Gangl, 2007) was used to calculate a Wilcoxon's singed rank test statistic and the Hodges-Lehmann point and interval estimates (DiPrete & Gangl, 2004). Table 4.A, for example, shows the results of the Rosenbaum bounds sensitivity analysis for PPVT Round 5 with the range of significance levels for the Wilcoxon's Singed Rank Statistic. It shows that the model became sensitive to hidden bias at $\Gamma = 2.15$ for PPVT Round 5, which is the critical test statistic value before exceeding the conventional 0.05 significance level. For the models with binary outcome variables (e.g., on-time grade progression), Stata's *mhbounds* programme (Becker & Caliendo, 2007) was used to calculate the Mantel-Haenszel test statistic (Mantel & Haenszel, 1959).

Table C. 1. Results of the Sensitivity Analysis (rbounds) for PPVT Round 5 Outcome: Range of Significance Levels for the Signed Rank Statistic

Gamma (Γ)	Minimum	Maximum
1	< .00001	< .00001
1.5	0	.000072
1.8	0	.003159
1.9	0	.007811
2	0	.016913
2.1	0	.032708
2.15	0	.043828
2.2	0	.057432
2.3	0	.092816

Note: (1) Gamma is log odds of differential assignment due to unobserved factors.

Table 4.B summarises the model's sensitivity to hidden bias across the outcome variables used in the present study. All matching models used kernel-based PSM with the Epanechikov kernel within the bandwidth at 0.11. As a results of Rosenbaum bounds sensitivity analysis, the model's sensitivity is at $\Gamma = 2.15$ for PPVT, $\Gamma = 1.1$ for math, $\Gamma = 1.45$ for language, $\Gamma = 2.2$ for the highest grade achieved, and $\Gamma = 2.5$ for on-time grade progression at age 15. For example, in interpreting the results of on-time grade progression at age 15, the test statistic value of 2.5 (Γ) [169] indicates that, in order for the 95 percent confidence interval of the model's ATT to include zero, an unobserved variable would need to cause the odds ratio of treatment assignment to differ between the treatment and comparison groups by a factor of 2.5 (Becker & Caliendo, 2007). Compared to the PPVT score, the highest grade achieved, and on-time grade progression, the results of the math and language scores were a relatively small value, indicating their high sensitivity to hidden bias; therefore, further analysis that controls for additional biases is warranted. As a reference point, in the logistic regression model used to generate the propensity score, the odds ratio for preschool attendance was the greatest for living in an urban area (OR=3.64) and being a first born (OR=3.62), followed by being in the high-achievement group (OR=2.28), living with parents with high educational aspirations (OR=1.43), and coming from the richest households (OR=1.05).

[169] The test statistic value *gamma* is a log odds of differential assignment due to unobserved factors.

Table C. 2. Results of Sensitivity Test

Outcome	Sensitivity	
	Pooled sample	Urban sample
Rosenbaum bounds sensitivity analysis		
PPVT – Round 3	1.8	1.05
PPVT – Round 4	1.8	1.45
PPVT – Round 5	2.15	1.75
Math – Round 3	2.2	2.25
Math – Round 4	1.0	1.0
Math – Round 5	1.1	1.05
Language - Round 3	1.25	1.0
Language - Round 4	1.0	1.0
Language - Round 5	1.45	1.95
Highest Grade – Round 5	2.2	1.7
Mantel-Haenszel bounds sensitivity analysis		
On-time progression – Round 4	1.8	1.2
On-time progression – Round 5	2.5	1.6

Note: (1) All models used the bandwidth at 0.11; (2) The 'Sensitivity' column presents the critical value at which the Rosenbaum or Mantel-Haenszel test statistics' significance level exceeds the conventional 0.05 level; (3) All matching models used kernel-based matching with the Epanechikov kernel.
Source: Young Lives Dataset Round 2 to Round 5, Young Lives

Appendix D. [Chapter 2] TABLES

Table D. 1. EGRA 2010: Preschool Attendance and Oral Reading Fluency (correct words per minute)

Variables	(1) ORF	(2) ORF	(3) ORF	(4) ORF
Preschool	2.46**	1.24	0.46	-0.39
	(1.22)	(1.44)	(1.55)	(1.94)
Age		1.81*	1.81*	1.13
		(1.02)	(0.98)	(0.73)
Age*Age		-0.04	-0.04	-0.02
		(0.04)	(0.03)	(0.03)
Female		-3.02***	-3.04***	-3.23***
		(1.03)	(1.04)	(0.85)
Same language at home			1.97	-1.91
			(2.77)	(2.14)
Reading materials at home			5.99***	5.96***
			(1.73)	(1.62)
Father's literacy			2.94**	2.18***
			(1.11)	(0.71)
Mother's literacy			-0.45	1.44*
			(1.00)	(0.79)
Region: Tigray	11.74***	11.64***	11.77***	-
	(2.70)	(2.76)	(2.81)	
Region: Amhara	15.87***	15.31***	14.92***	-
	(2.91)	(3.06)	(2.94)	
Region: Oromia	19.46***	16.91***	17.34***	-
	(3.48)	(3.16)	(3.04)	
Region: Somali	18.82***	15.23***	16.40***	-
	(6.59)	(4.95)	(4.29)	
Region: SNNP (ref)	-	-	-	-
Grade 3	8.20***	7.24***	7.06***	7.46***
	(1.04)	(1.16)	(1.17)	(0.92)
Grade 2 (ref)	-	-	-	-
Rural	-	-8.09***	-7.12***	-
		(2.17)	(1.91)	
Constant	3.85*	0.17	-5.09	8.63*
	(2.01)	(6.21)	(5.84)	(4.99)
Observations	9,121	9,121	9,121	9,121
R-squared	0.15	0.18	0.20	0.10
Number of schools				237

Note: (1) Models 1, 2, and 3 account for controls as indicated and include sampling weight; (2) Model 4 uses school fixed effects and includes sampling weight; (3) EGRA 2010: linearised standard errors (from svy command) in parentheses; (4) EGRA 2016: robust standard errors, clustered at school level, in parentheses
*** p<0.01, ** p<0.05, *p<0.1
Source: EGRA Dataset 2010, 2016, USAID

Table D. 2. EGRA 2016: Preschool Attendance and Oral Reading Fluency (correct words per minute)

Variables	(1) ORF	(2) ORF	(3) ORF	(4) ORF
Preschool	5.36***	5.45***	4.15***	2.48***
	(0.86)	(0.87)	(0.78)	(0.61)
Age		0.77	0.63	0.90
		(1.58)	(1.59)	(1.46)
Age*Age		0.02	0.03	0.01
		(0.07)	(0.07)	(0.07)
Female		2.39***	2.33***	2.29***
		(0.61)	(0.59)	(0.60)
Same language at home			0.59	1.81*
			(1.84)	(1.08)
Reading materials at home			5.29***	4.82***
			(0.58)	(0.48)
Father's literacy			3.10***	2.77***
			(0.53)	(0.50)
Mother's literacy			0.64	0.27
			(0.54)	(0.53)
Region: Tigray	0.57	1.26	0.39	-
	(2.05)	(1.91)	(1.82)	
Region: Amhara	15.31***	16.12***	15.45***	-
	(2.09)	(2.03)	(1.95)	
Region: Oromia	-3.82**	-3.27*	-2.99	-
	(1.83)	(1.86)	(1.85)	
Region: Somali	-8.05***	-7.29***	-7.26***	-
	(1.92)	(2.02)	(1.89)	
Region: SNNP (ref)	-	-	-	-
Grade 3	10.96***	9.81***	9.41***	9.29***
	(0.63)	(0.71)	(0.74)	(0.71)
Grade 2 (ref)	-	-	-	-
Rural	-	-6.21***	-6.17***	-
		(1.64)	(1.56)	
Constant	12.66***	6.88	2.89	2.62
	(1.50)	(9.05)	(8.69)	(7.86)
Observations	8,332	8,332	8,332	8,332
R-squared	0.22	0.24	0.26	0.11
Number of schools				225

Note: (1) Models 1, 2, and 3 account for controls as indicated and include sampling weight; (2) Model 4 uses school fixed effects and includes sampling weight; (3) EGRA 2010: linearised standard errors (from svy command) in parentheses; (4) EGRA 2016: robust standard errors, clustered at school level, in parentheses

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

Source: EGRA Dataset 2010, 2016, USAID

Table D. 3. EGRA 2010: Preschool Attendance and Non-Reader (odds ratio)

Variables	(1) Non-reader	(2) Non-reader	(3) Non-reader	(4) Non-reader
Preschool	0.94	0.99	1.06	1.28***
	(0.13)	(0.16)	(0.17)	(0.04)
Age		0.75***	0.74***	0.81***
		(0.08)	(0.08)	(0.02)
Age*Age		1.01**	1.01**	1.00***
		(0.00)	(0.00)	(0.00)
Female		1.36***	1.37***	1.66***
		(0.15)	(0.16)	(0.04)
Same language at home			0.57*	0.97
			(0.17)	(0.05)
Reading materials at home			0.55***	0.47***
			(0.07)	(0.01)
Father's literacy			0.60***	0.55***
			(0.07)	(0.01)
Mother's literacy			1.15	0.82***
			(0.11)	(0.02)
Region: Tigray	0.16***	0.16***	0.13***	-
	(0.04)	(0.04)	(0.04)	
Region: Amhara	0.16***	0.17***	0.16***	-
	(0.05)	(0.05)	(0.05)	
Region: Oromia	0.27***	0.31***	0.27***	-
	(0.07)	(0.08)	(0.07)	
Region: Somali	0.18***	0.23***	0.17***	-
	(0.07)	(0.09)	(0.06)	
Region: SNNP (ref)	-	-	-	-
Grade 3	0.46***	0.51***	0.51***	0.42***
	(0.04)	(0.05)	(0.05)	(0.01)
Grade 2 (ref)	-	-	-	-
Rural	-	1.53*	1.41	-
		(0.35)	(0.32)	
Constant	2.45***	9.03***	25.84***	5.43***
	(0.54)	(6.35)	(20.59)	(1.87)
Observations	9,121	9,121	9,121	8,812
Pseudo R2	0.10	0.11	0.13	-
Number of schools				229

Note: (1) All Models 1 to 4 include sampling weight; (2) Model 4 use school fixed effects; (3) EGRA 2010: linearised standard errors (from *svy* command) in parentheses; (4) EGRA 2016: robust standard errors, clustered at school level, in parentheses; (5) In Model 4, the sample of 8 groups (309 obs.) in non-reader and 10 groups (390 obs.) in proficient reader dropped as multiple positive or negative outcomes within groups encountered.
*** p<0.01, ** p<0.05, *p<0.1
Source: EGRA Dataset 2010, 2016, USAID

Table D. 4. EGRA 2016: Preschool Attendance and Non-Reader (odds ratio)

Variables	(1) Non-reader	(2) Non-reader	(3) Non-reader	(4) Non-reader
Preschool	0.46***	0.42***	0.49***	0.77***
	(0.05)	(0.04)	(0.05)	(0.01)
Age		0.48***	0.50**	0.64***
		(0.12)	(0.13)	(0.01)
Age*Age		1.03**	1.02*	1.01***
		(0.01)	(0.01)	(0.00)
Female		0.89	0.89	0.70***
		(0.08)	(0.07)	(0.00)
Same language at home			0.78	0.74***
			(0.18)	(0.01)
Reading materials at home			0.47***	0.49***
			(0.05)	(0.00)
Father's literacy			0.61***	0.68***
			(0.06)	(0.01)
Mother's literacy			0.94	1.10***
			(0.09)	(0.01)
Region: Tigray	0.44***	0.41***	0.45***	-
	(0.10)	(0.10)	(0.10)	
Region: Amhara	0.07***	0.07***	0.07***	-
	(0.02)	(0.02)	(0.02)	
Region: Oromia	0.95	0.91	0.88	-
	(0.20)	(0.19)	(0.18)	
Region: Somali	1.84***	1.71**	1.70**	-
	(0.41)	(0.41)	(0.43)	
Region: SNNP (ref)	-	-	-	-
Grade 3	0.39***	0.47***	0.50***	0.46***
	(0.03)	(0.04)	(0.05)	(0.00)
Grade 2 (ref)	-	-	-	-
Rural	-	1.52**	1.59**	-
		(0.32)	(0.34)	
Constant	1.01	70.75***	132.12***	1.68***
	(0.17)	(94.85)	(186.38)	(0.23)
Observations	8,332	8,332	8,332	7,461
Pseudo R2	0.16	0.18	0.20	-
Number of schools				202

Note: (1) All Models 1 to 4 include sampling weight; (2) Model 4 use school fixed effects; (3) EGRA 2010: linearised standard errors (from *svy* command) in parentheses; (4) EGRA 2016: robust standard errors, clustered at school level, in parentheses; (5) In Model 4, the sample of 23 groups (871 obs.) in non-reader and 11 groups (312 obs.) dropped as multiple positive or negative outcomes within groups encountered.

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

Source: EGRA Dataset 2010, 2016, USAID

Table D. 5. EGRA 2010: Preschool Attendance and Proficient Reader (odds ratio)

Variables	(1) Proficient reader	(2) Proficient reader	(3) Proficient reader	(4) Proficient reader
Preschool	1.23*	1.11	1.05	1.19***
	(0.13)	(0.11)	(0.10)	(0.04)
Age		1.27**	1.27**	1.57***
		(0.12)	(0.12)	(0.05)
Age*Age		0.99**	0.99**	0.98***
		(0.00)	(0.00)	(0.00)
Female		0.78**	0.78**	0.72***
		(0.08)	(0.08)	(0.01)
Same language at home			1.27	0.85***
			(0.34)	(0.04)
Reading materials at home			1.59***	1.61***
			(0.17)	(0.04)
Father’s literacy			1.31**	1.29***
			(0.15)	(0.03)
Mother’s literacy			0.91	1.14***
			(0.10)	(0.03)
Region: Tigray	4.27***	4.30***	4.55***	-
	(1.40)	(1.45)	(1.44)	
Region: Amhara	2.74***	2.65***	2.59***	-
	(0.89)	(0.88)	(0.84)	
Region: Oromia	5.16***	4.33***	4.60***	-
	(1.84)	(1.49)	(1.54)	
Region: Somali	5.99***	4.64***	5.28***	-
	(2.44)	(1.80)	(1.92)	
Region: SNNP (ref)	-	-	-	-
Grade 3	1.57***	1.47***	1.45***	1.65***
	(0.14)	(0.13)	(0.13)	(0.04)
Grade 2 (ref)	-	-	-	-
Rural	-	0.54***	0.58***	-
		(0.12)	(0.12)	
Constant	0.15***	0.06***	0.04***	3.29***
	(0.04)	(0.05)	(0.03)	(0.86)
Observations	9,121	9,121	9,121	8,731
Pseudo R2	0.07	0.08	0.09	-
Number of schools				227

Note: (1) All Models 1 to 4 include sampling weight; (2) Model 4 use school fixed effects; (3) EGRA 2010: linearised standard errors (from *svy* command) in parentheses; (4) EGRA 2016: robust standard errors, clustered at school level, in parentheses; (5) In Model 4, the sample of 8 groups (309 obs.) in non-reader and 10 groups (390 obs.) in proficient reader dropped as multiple positive or negative outcomes within groups encountered.

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

Source: EGRA Dataset 2010, 2016, USAID

Table D. 6. EGRA 2016: Preschool Attendance and Proficient Reader (odds ratio)

Variables	(1) Proficient reader	(2) Proficient reader	(3) Proficient reader	(4) Proficient reader
Preschool	1.57***	1.58***	1.38***	1.07***
	(0.15)	(0.15)	(0.12)	(0.01)
Age		1.29	1.28	0.99
		(0.25)	(0.25)	(0.02)
Age*Age		0.99	0.99	1.00***
		(0.01)	(0.01)	(0.00)
Female		1.35***	1.35***	1.56***
		(0.10)	(0.11)	(0.01)
Same language at home			0.93	0.95**
			(0.18)	(0.02)
Reading materials at home			1.70***	1.72***
			(0.12)	(0.01)
Father's literacy			1.37***	1.32***
			(0.09)	(0.01)
Mother's literacy			1.11	1.06***
			(0.08)	(0.01)
Region: Tigray	1.10	1.20	1.10	-
	(0.26)	(0.27)	(0.25)	
Region: Amhara	1.76**	1.96***	1.86***	-
	(0.40)	(0.44)	(0.41)	
Region: Oromia	0.49***	0.52***	0.52***	-
	(0.11)	(0.12)	(0.12)	
Region: Somali	0.35***	0.37***	0.35***	-
	(0.09)	(0.10)	(0.09)	
Region: SNNP (ref)	-	-	-	-
Grade 3	1.96***	1.76***	1.73***	1.75***
	(0.14)	(0.13)	(0.13)	(0.01)
Grade 2 (ref)	-	-	-	-
Rural		0.51***	0.50***	-
		(0.10)	(0.10)	
Constant	0.44***	0.10**	0.07**	1.50***
	(0.08)	(0.11)	(0.08)	(0.19)
Observations	8,332	8,332	8,332	8,020
Pseudo R2	0.08	0.10	0.11	
Number of schools				214

Note: (1) All Models 1 to 4 include sampling weight; (2) Model 4 use school fixed effects; (3) EGRA 2010: linearised standard errors (from *svy* command) in parentheses; (4) EGRA 2016: robust standard errors, clustered at school level, in parentheses; (5) In Model 4, the sample of 23 groups (871 obs.) in non-reader and 11 groups (312 obs.) dropped as multiple positive or negative outcomes within groups encountered.

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

Source: EGRA Dataset 2010, 2016, USAID

Table D. 7. EGRA 2010: Preschool Attendance and EGRA 6 Sub-Tasks (SD, effect size)

Variables	(1) ORF	(2) Letter Sounds	(3) Familiar Words	(4) Invented Words	(5) Reading Compre.	(6) Listening Compre.
Preschool	0.02	0.11**	0.08	0.11**	0.05	0.01
	(0.07)	(0.05)	(0.07)	(0.06)	(0.07)	(0.06)
Age	0.09*	0.04	0.03	0.06	0.10**	0.13***
	(0.05)	(0.04)	(0.04)	(0.04)	(0.04)	(0.05)
Age*Age	-0.00	-0.00	-0.00	-0.00	-0.00*	-0.00*
	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)
Female	-0.15***	-0.17***	-0.15***	-0.11**	-0.08	-0.04
	(0.05)	(0.04)	(0.05)	(0.05)	(0.05)	(0.03)
Same language	0.09	0.29**	0.11	0.09	0.07	0.31***
	(0.13)	(0.12)	(0.13)	(0.14)	(0.15)	(0.08)
Reading M at home	0.29***	0.20***	0.25***	0.23***	0.25**	0.20***
	(0.08)	(0.06)	(0.07)	(0.06)	(0.10)	(0.05)
Father’s literacy	0.14**	0.17***	0.18***	0.18***	0.13**	0.16***
	(0.05)	(0.06)	(0.05)	(0.05)	(0.06)	(0.04)
Mother’s literacy	-0.02	-0.02	-0.02	0.01	-0.04	0.04
	(0.05)	(0.04)	(0.04)	(0.04)	(0.05)	(0.04)
Region: Tigray	0.56***	0.21	0.90***	0.53***	0.36***	0.30***
	(0.13)	(0.19)	(0.16)	(0.14)	(0.09)	(0.10)
Region: Amhara	0.71***	0.47***	0.79***	0.60***	0.53***	0.13
	(0.14)	(0.16)	(0.15)	(0.12)	(0.12)	(0.16)
Region: Oromia	0.83***	0.50***	0.53***	0.35***	0.86***	1.36***
	(0.15)	(0.14)	(0.13)	(0.10)	(0.14)	(0.11)
Region: Somali	0.79***	0.48***	0.40***	0.75***	0.58***	0.07
	(0.21)	(0.15)	(0.10)	(0.13)	(0.15)	(0.15)
Region: SNNP (ref)	-	-	-	-	-	-
Grade 3	0.34***	0.34***	0.39***	0.32***	0.36***	0.12***
	(0.06)	(0.05)	(0.05)	(0.04)	(0.05)	(0.04)
Grade 2 (ref)	-	-	-	-	-	-
Rural	-0.34***	-0.30***	-0.28***	-0.34***	-0.31***	-0.17**
	(0.09)	(0.07)	(0.10)	(0.08)	(0.10)	(0.07)
Constant	-1.26***	-0.94***	-0.98***	-0.97***	-1.25***	-1.78***
	(0.28)	(0.30)	(0.33)	(0.31)	(0.30)	(0.30)
Observations	9,121	9,119	9,119	9,118	9,121	9,116
R-squared	0.20	0.13	0.18	0.15	0.19	0.35

Note: (1) All models 1 to 6 include sampling weight; (2) Linearised standard errors (from svy command) in parentheses.
*** p<0.01, ** p<0.05, *p<0.1
Source: EGRA Dataset 2010, 2016, USAID

Table D. 8. EGRA 2016: Preschool Attendance and EGRA 6 Sub-Tasks (SD, effect size)

Variables	(1) ORF	(2) Letter Sounds	(3) Familiar Words	(4) Invented Words	(5) Reading Compre.	(6) Listening Compre.
Preschool	0.20***	0.20***	0.21***	0.19***	0.17***	0.07**
	(0.03)	(0.03)	(0.03)	(0.03)	(0.03)	(0.03)
Age	0.03	0.12*	0.03	0.11	0.01	0.19**
	(0.07)	(0.07)	(0.07)	(0.07)	(0.07)	(0.08)
Age*Age	0.00	-0.00	0.00	-0.00	0.00	-0.01*
	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)
Female	0.11***	0.08***	0.09***	0.09***	0.15***	-0.08***
	(0.02)	(0.03)	(0.02)	(0.03)	(0.03)	(0.03)
Same language	0.03	-0.01	-0.01	-0.07	0.05	0.20***
	(0.07)	(0.07)	(0.06)	(0.08)	(0.07)	(0.07)
Reading M at home	0.26***	0.27***	0.26***	0.27***	0.23***	0.17***
	(0.03)	(0.03)	(0.03)	(0.03)	(0.03)	(0.03)
Father's literacy	0.15***	0.22***	0.17***	0.19***	0.12***	0.09***
	(0.03)	(0.03)	(0.03)	(0.03)	(0.03)	(0.03)
Mother's literacy	0.03	0.05*	0.03	0.04	0.04	-0.01
	(0.03)	(0.03)	(0.03)	(0.03)	(0.03)	(0.03)
Region: Tigray	0.02	-0.72***	0.50***	0.00	-0.18***	-0.31***
	(0.04)	(0.04)	(0.04)	(0.04)	(0.04)	(0.03)
Region: Amhara	0.75***	-0.10**	0.85***	0.59***	0.44***	-0.55***
	(0.04)	(0.04)	(0.04)	(0.04)	(0.04)	(0.04)
Region: Oromia	-0.14***	-0.33***	-0.15***	-0.55***	-0.10**	0.11***
	(0.04)	(0.04)	(0.04)	(0.04)	(0.04)	(0.03)
Region: Somali	-0.35***	-0.78***	-0.37***	-0.38***	-0.31***	-0.76***
	(0.04)	(0.05)	(0.04)	(0.06)	(0.05)	(0.05)
Region: SNNP (ref)	-	-	-	-	-	-
Grade 3	0.45***	0.35***	0.43***	0.35***	0.48***	0.12***
	(0.03)	(0.03)	(0.03)	(0.03)	(0.03)	(0.03)
Grade 2 (ref)	-	-	-	-	-	-
Rural	-0.30***	-0.33***	-0.38***	-0.26***	-0.30***	-0.20***
	(0.03)	(0.04)	(0.04)	(0.04)	(0.04)	(0.03)
Constant	-0.81**	-0.71*	-0.78**	-0.96**	-0.69*	-1.16***
	(0.36)	(0.37)	(0.36)	(0.38)	(0.38)	(0.42)
Observations	8,332	8,332	8,332	8,332	8,332	8,332
R-squared	0.26	0.23	0.32	0.26	0.19	0.16

Note: (1) All models 1 to 6 include sampling weight; (2) Robust standard errors in parentheses.
*** p<0.01, ** p<0.05, *p<0.1
Source: EGRA Dataset 2010, 2016, USAID

Table D. 9. EGRA 2016: Preschool Attendance and EGRA 6 Sub-Tasks with School Fixed Effects Model (SD, effect size)

Variables	(1) ORF	(2) Letter Sounds	(3) Familiar Words	(4) Invented Words	(5) Reading Compre.	(6) Listening Compre.
Preschool	0.12***	0.11***	0.11***	0.11***	0.10***	0.04
	(0.03)	(0.03)	(0.03)	(0.03)	(0.03)	(0.03)
Age	0.04	0.15*	0.08	0.15*	0.05	0.24***
	(0.07)	(0.09)	(0.07)	(0.08)	(0.08)	(0.06)
Age*Age	0.00	-0.00	-0.00	-0.00	0.00	-0.01***
	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)
Female	0.11***	0.07**	0.09***	0.08**	0.15***	-0.08***
	(0.03)	(0.03)	(0.03)	(0.03)	(0.03)	(0.03)
Same language	0.09*	0.10	0.08	0.06	0.10*	0.17**
	(0.05)	(0.08)	(0.05)	(0.06)	(0.06)	(0.08)
Reading M at home	0.23***	0.25***	0.23***	0.26***	0.21***	0.14***
	(0.02)	(0.03)	(0.02)	(0.03)	(0.03)	(0.03)
Father's literacy	0.13***	0.19***	0.15***	0.15***	0.11***	0.08***
	(0.02)	(0.03)	(0.03)	(0.02)	(0.03)	(0.03)
Mother's literacy	0.01	0.01	0.01	0.01	0.03	-0.04*
	(0.03)	(0.02)	(0.02)	(0.03)	(0.03)	(0.02)
Grade 3	0.42***	0.35***	0.40***	0.34***	0.45***	0.13***
	(0.03)	(0.03)	(0.03)	(0.03)	(0.03)	(0.03)
Grade 2 (ref)	-	-	-	-	-	-
Rural	-	-	-	-	-	-
Constant	-0.73***	-1.24***	-0.90***	-0.94***	-0.75***	-1.88***
	(0.26)	(0.27)	(0.25)	(0.26)	(0.28)	(0.31)
Observations	8,332	8,332	8,332	8,332	8,332	8,332
R-squared	0.11	0.11	0.12	0.09	0.11	0.04
Number of Schools	225	225	225	225	225	225

Note: (1) All models 1 to 6 include sampling weight; (2) Robust standard errors in parentheses. (3) School fixed effects model does not account for five regional variables.
*** p<0.01, ** p<0.05, *p<0.1
Source: EGRA Dataset 2010, 2016, USAID

Table D. 10. EGRA 2016: Results of SEM with Observable School Characteristics

	(1) ORF	(2) LS	(3) FW	(4) IW	(5) RC	(6) LC
	Correct words/letters per minute				% of correct answers	
***Path Coefficients* (S.E.)**						
Preschool Attendance → School Structural Quality (Environments) → EGRA Outcome						
Total effects	2.40***	8.54***	2.82***	2.30***	3.32***	4.22***
	(0.46)	(0.70)	(0.46)	(0.33)	(0.60)	(0.62)
Direct effects	2.00***	8.38***	2.21***	1.90***	2.97***	4.23***
	(0.46)	(0.70)	(0.45)	(0.33)	(0.60)	(0.63)
Indirect effects	0.40***	0.15***	0.61***	0.40***	0.36***	*-0.01*
	(0.08)	(0.10)	(0.10)	(0.06)	(0.09)	*(0.06)*
% of total effect mediated	16.7	1.9	21.6	17.4	10.5	-0.2
% of total effect unmediated	83.3	98.1	78.4	82.6	89.5	100.2
Ratio of indirect effect to direct effect (R*m*)	0.20	0.02	0.28	0.21	0.12	0.00
Model Fit						
Chi-Square (*DF*)	722.48(3)	722.48(3)	722.48(3)	722.48(3)	722.48(3)	722.48(3)
CFI	0.73	0.74	0.75	0.72	0.71	0.60
RMSEA	0.17	0.17	0.17	0.17	0.17	0.17
SRMR	0.03	0.03	0.03	0.03	0.03	0.03
R-Square	0.18	0.20	0.19	0.16	0.18	0.12

Note: (1) The resulting structural coefficients (standardised regression coefficients) describe the direct and indirect effects. (2) Standard errors are in parentheses; (3) ORF oral reading fluency; LS Letter sound; FW familiar word recognition; IW invented word recognition; RC reading comprehension; LC listening comprehension; (4) DF: Degree of Freedom; (5) CFI: Comparative Fit Index; (6) RMSEA: Root Mean Square Error of Approximation; (7) SRMR: Standardised Root Mean Square Residual.
*** p<0.01, ** p<0.05, *p<0.1
Source: EGRA Dataset 2010, 2016, USAID

Appendix E. [Chapter 4] TABLES

Table E. 1. Young Lives: Relation between Preschool and Academic Achievement, Full OLS Model (Pooled sample) (standardized scores)

Variables	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)
							Language			Supplementary
	PPVT			Math			Mother Tongue		English	Oral Reading Fluency
Round (Age)	Round 3 (age 8)	Round 4 (age 12)	Round 5 (age 15)	Round 3 (age 8)	Round 4 (age 12)	Round 5 (age 15)	Round 3 (age 8)	Round 4 (age 12)	Round 5 (age 15)	Round 3 (age 8)
Preschool attendance	0.43***	0.34***	0.24***	0.38***	0.02	0.04	0.15*	0.12	0.24**	0.35**
	(0.12)	(0.08)	(0.08)	(0.11)	(0.10)	(0.05)	(0.10)	(0.11)	(0.10)	(0.15)
Private Spending on education (vs. high)										
Low	-0.15***	-0.10*	-0.13**	-0.24***	-0.13**	-0.15**	-0.22**	-0.13**	-0.16**	-0.24**
	(0.04)	(0.05)	(0.05)	(0.06)	(0.05)	(0.07)	(0.09)	(0.05)	(0.07)	(0.10)
Middle	0.00	0.01	-0.02	-0.02	0.03	0.00	-0.07	-0.04	-0.06*	-0.10
	(0.04)	(0.05)	(0.05)	(0.07)	(0.05)	(0.05)	(0.07)	(0.04)	(0.03)	(0.06)
Parental aspiration	0.18***	0.18***	0.12*	0.10	0.08*	0.09*	0.11*	0.09	0.17***	0.03
	(0.06)	(0.05)	(0.06)	(0.06)	(0.04)	(0.05)	(0.06)	(0.06)	(0.05)	(0.04)
Age 14 (vs. Age 15)	-0.21***	-0.03	-0.03	-0.15***	-0.08	0.04	-0.10	-0.10*	-0.02	-0.10*
	(0.04)	(0.06)	(0.05)	(0.04)	(0.06)	(0.05)	(0.06)	(0.05)	(0.05)	(0.05)
Female (vs. Male)	0.01	0.01	-0.03	0.00	0.06	-0.02	0.06	0.13*	0.10*	0.07*
	(0.05)	(0.05)	(0.05)	(0.05)	(0.06)	(0.05)	(0.05)	(0.07)	(0.05)	(0.04)
PPVT at age 5 (vs. Low)										
Middle	0.06	0.20***	0.21***	0.04	0.08	0.01	0.03	0.12*	0.13**	-0.00
	(0.05)	(0.06)	(0.05)	(0.06)	(0.08)	(0.05)	(0.05)	(0.06)	(0.06)	(0.05)
High	0.12**	0.25***	0.22***	0.21**	0.21***	0.12**	0.12	0.21***	0.18***	0.02
	(0.06)	(0.07)	(0.05)	(0.08)	(0.06)	(0.06)	(0.08)	(0.06)	(0.04)	(0.05)
CDA-Q(Math) at age 5 (vs. High)										
Low	-0.24***	-0.15*	-0.11	-0.16**	-0.17**	-0.19**	-0.13	-0.13*	-0.21***	-0.24***
	(0.06)	(0.08)	(0.07)	(0.07)	(0.08)	(0.07)	(0.08)	(0.07)	(0.06)	(0.08)
Middle	-0.19***	-0.11**	-0.11**	-0.15**	-0.11	-0.16***	-0.12**	-0.07	-0.14***	-0.25***
	(0.04)	(0.04)	(0.05)	(0.06)	(0.08)	(0.05)	(0.05)	(0.05)	(0.04)	(0.06)
Height-for-age z-score at age 5 (vs. High)										
Low	-0.18**	-0.17**	-0.09	-0.16***	-0.09**	-0.12**	-0.23***	-0.13**	-0.01	-0.12*
	(0.07)	(0.07)	(0.05)	(0.04)	(0.03)	(0.05)	(0.07)	(0.05)	(0.05)	(0.06)
Middle	-0.20***	-0.10	-0.04	-0.11**	-0.07	-0.11*	-0.15**	-0.12**	-0.04	-0.05
	(0.06)	(0.06)	(0.05)	(0.04)	(0.04)	(0.06)	(0.06)	(0.05)	(0.05)	(0.05)

Health prob. at age 5(vs. No)	-0.14*	-0.08	-0.11	-0.06	-0.04	0.02	-0.03	-0.05	0.04	-0.01
	(0.07)	(0.07)	(0.07)	(0.06)	(0.06)	(0.08)	(0.07)	(0.10)	(0.08)	(0.07)
Father's highest education level (vs. Secondary and above)										
No education	-0.07	-0.09	-0.07	-0.31***	-0.33***	-0.31***	-0.24**	-0.19**	-0.31***	-0.16
	(0.09)	(0.06)	(0.06)	(0.07)	(0.07)	(0.06)	(0.10)	(0.08)	(0.06)	(0.10)
Primary education	-0.10	-0.05	-0.10**	-0.20***	-0.22***	-0.19**	-0.09	-0.12*	-0.18**	-0.08
	(0.07)	(0.05)	(0.05)	(0.05)	(0.07)	(0.07)	(0.08)	(0.07)	(0.07)	(0.08)
Caregiver's highest education level (vs. Secondary and above)										
No education	-0.18*	-0.07	-0.06	-0.05	-0.13	0.02	0.01	-0.10	-0.01	0.03
	(0.10)	(0.10)	(0.10)	(0.06)	(0.10)	(0.07)	(0.10)	(0.12)	(0.10)	(0.09)
Primary education	-0.12	-0.03	-0.01	-0.01	-0.12	-0.06	0.02	0.02	0.04	0.06
	(0.09)	(0.06)	(0.06)	(0.07)	(0.11)	(0.08)	(0.09)	(0.12)	(0.09)	(0.08)
Household size (> 6)	-0.10**	-0.17**	-0.11**	-0.15***	-0.13**	-0.03	-0.11*	-0.14*	-0.03	-0.12
	(0.04)	(0.06)	(0.05)	(0.03)	(0.05)	(0.06)	(0.06)	(0.07)	(0.05)	(0.07)
First born	-0.00	0.05	0.06	-0.05	-0.01	0.03	0.06	0.05	0.17***	0.08
	(0.04)	(0.03)	(0.05)	(0.05)	(0.06)	(0.08)	(0.07)	(0.07)	(0.05)	(0.06)
Same language home/school	-0.13	-0.03	-0.01	-0.08	0.02	0.08	0.02	-0.04	-0.07	-0.04
	(0.08)	(0.07)	(0.07)	(0.09)	(0.05)	(0.07)	(0.08)	(0.06)	(0.06)	(0.09)
Wealth quintile (vs. Quintile 4)										
Quintile 1	-0.07	-0.19	-0.44***	-0.24***	-0.13	-0.16	-0.21**	-0.22**	-0.11	-0.09
	(0.08)	(0.13)	(0.12)	(0.08)	(0.10)	(0.10)	(0.08)	(0.10)	(0.09)	(0.08)
Quintile 2	-0.06	-0.12	-0.25**	-0.14**	-0.07	-0.11	-0.21**	-0.08	-0.05	-0.06
	(0.08)	(0.10)	(0.10)	(0.06)	(0.09)	(0.09)	(0.08)	(0.09)	(0.08)	(0.07)
Quintile 3	-0.02	0.02	-0.16**	-0.09	-0.04	-0.08	-0.08	-0.07	-0.09	-0.04
	(0.05)	(0.08)	(0.06)	(0.05)	(0.06)	(0.07)	(0.06)	(0.07)	(0.07)	(0.07)
Quintile 5	0.23***	0.08	0.10*	0.16**	0.14*	0.19	0.04	0.06	-0.03	0.19**
	(0.08)	(0.06)	(0.05)	(0.07)	(0.07)	(0.14)	(0.05)	(0.10)	(0.06)	(0.08)
Region (vs. Addis Ababa)										
Living in Tigray	-0.04	-0.18	-0.31	0.16	-0.41***	-1.01***	0.19	-0.08	-0.39**	-0.38
	(0.30)	(0.24)	(0.20)	(0.18)	(0.13)	(0.16)	(0.25)	(0.10)	(0.17)	(0.23)
Living in Amhara	-0.38**	-0.37	-0.14	0.00	-0.10	-0.49***	-0.12	0.04	0.06	-0.51***
	(0.18)	(0.24)	(0.17)	(0.20)	(0.14)	(0.08)	(0.18)	(0.11)	(0.10)	(0.17)
Living in Oromia	-0.55***	0.24	0.03	-0.34*	-0.28**	-0.77***	-0.32**	-0.74***	-0.65***	-0.79***
	(0.15)	(0.21)	(0.15)	(0.17)	(0.13)	(0.09)	(0.14)	(0.11)	(0.10)	(0.11)
Living in SNNP	-0.15	-0.43***	-0.44***	-0.05	-0.48***	-0.76***	-0.23	-0.42**	-0.54***	-0.70***
	(0.26)	(0.11)	(0.14)	(0.23)	(0.14)	(0.08)	(0.17)	(0.18)	(0.12)	(0.18)
Child's ethnicity (vs. SNNP)										
Ethnicity 1: Others	-0.17	-0.01	-0.32	0.04	0.15*	0.08	0.10	0.19*	0.06	0.37***
	(0.24)	(0.18)	(0.19)	(0.16)	(0.08)	(0.08)	(0.13)	(0.10)	(0.13)	(0.07)
Ethnicity 3: Oromiffa	0.01	-0.13	-0.29	0.01	-0.09	-0.02	0.12	0.14	0.02	0.31*

	(0.23)	(0.15)	(0.17)	(0.17)	(0.08)	(0.09)	(0.14)	(0.12)	(0.13)	(0.16)
Ethnicity 4: Tigrian	0.02	0.10	-0.09	0.19	0.13	0.32*	0.42	0.11	0.17	0.33
	(0.27)	(0.18)	(0.17)	(0.22)	(0.09)	(0.17)	(0.27)	(0.16)	(0.16)	(0.33)
Ethnicity 5: Amhara	0.16	0.04	-0.20	0.09	-0.04	0.01	0.18	0.11	0.12	0.41***
	(0.21)	(0.15)	(0.18)	(0.13)	(0.05)	(0.11)	(0.13)	(0.10)	(0.12)	(0.08)
Living in urban	0.13	0.44	0.46**	0.35***	0.46***	0.21***	0.34***	0.34***	0.29**	0.31***
	(0.17)	(0.27)	(0.20)	(0.08)	(0.08)	(0.06)	(0.10)	(0.06)	(0.11)	(0.07)
Constant	0.59	0.02	0.37	0.30	0.41*	0.73***	0.09	0.20	0.21	0.34
	(0.36)	(0.39)	(0.35)	(0.28)	(0.21)	(0.17)	(0.22)	(0.27)	(0.26)	(0.23)
Observations	1,447	1,417	1,447	1,417	1,354	1,447	1,444	1,320	1,447	1,390
R-squared	0.46	0.46	0.46	0.47	0.32	0.33	0.31	0.33	0.35	0.46

Note: (1) Preschool coefficients and ATTs are based on the standardised score (z-score) of each test; (2) robust standard errors, clustered at community level, in parentheses; (3) Ethnicity in SNNP includes Hadiva, Sidama, Wolayta; and others includes Agew, Gurage, Kambata.
*** p<0.01, ** p<0.05, *p<0.1
Source: Young Lives Dataset Round 2 to Round 5, Young Lives

Table E. 2. Young Lives: Relation between Preschool and Academic Achievement, Full OLS Model (Urban sample) (standardized scores)

Variables	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)
							Language			Supplementary
		PPVT			Math		Mother Tongue		English	Oral Reading Fluency
Round	Round 3	Round 4	Round 5	Round 3	Round 4	Round 5	Round 3	Round 4	Round 5	Round 3
(Age)	(age 8)	(age 12)	(age 15)	(age 8)	(age 12)	(age 15)	(age 8)	(age 12)	(age 15)	(age 8)
Preschool attendance	0.24*	0.30**	0.30***	0.27***	0.03	0.05	0.06	0.01	0.23**	0.27**
	(0.11)	(0.10)	(0.06)	(0.07)	(0.16)	(0.07)	(0.13)	(0.08)	(0.09)	(0.15)
Private Spending on education (vs. high)										
Low	-0.11	0.10	-0.09	-0.19	-0.06	-0.09	-0.13	-0.00	0.09	-0.13
	(0.07)	(0.12)	(0.09)	(0.11)	(0.06)	(0.07)	(0.15)	(0.12)	(0.10)	(0.15)
Middle	-0.08	0.08	-0.06	-0.13	-0.08	-0.07	-0.08	-0.02	-0.02	-0.01
	(0.06)	(0.13)	(0.08)	(0.10)	(0.09)	(0.09)	(0.12)	(0.08)	(0.06)	(0.08)
Parental aspiration	0.12	-0.02	-0.00	0.14	0.08	0.01	0.16*	0.17	0.17	-0.01
	(0.11)	(0.12)	(0.08)	(0.13)	(0.10)	(0.15)	(0.08)	(0.10)	(0.12)	(0.04)
Age 14 (vs. Age 15)	-0.24***	-0.09	0.07	-0.16*	-0.14	0.05	-0.16	-0.10	-0.05	-0.16
	(0.06)	(0.08)	(0.08)	(0.07)	(0.12)	(0.10)	(0.09)	(0.10)	(0.09)	(0.09)
Female (vs. Male)	0.00	0.03	-0.10	0.07	0.13	0.08	0.11	0.14	0.16	0.13**
	(0.08)	(0.09)	(0.08)	(0.07)	(0.11)	(0.09)	(0.08)	(0.13)	(0.09)	(0.05)
PPVT at age 5 (vs. High)										
Middle	-0.09	-0.21	-0.30***	-0.28***	-0.24**	-0.12	-0.14	-0.09	-0.11*	-0.07
	(0.10)	(0.14)	(0.09)	(0.08)	(0.08)	(0.09)	(0.09)	(0.07)	(0.06)	(0.09)
High	-0.13	-0.23*	-0.21*	-0.22*	-0.07	-0.20*	-0.24*	-0.04	-0.15	-0.18
	(0.12)	(0.12)	(0.11)	(0.12)	(0.14)	(0.09)	(0.11)	(0.10)	(0.09)	(0.11)
CDA-Q(Math) at age 5 (vs. High)										
Low	-0.31***	-0.30**	-0.10	-0.13	-0.11	-0.24*	-0.22	-0.23	-0.22**	-0.37***
	(0.08)	(0.13)	(0.09)	(0.12)	(0.12)	(0.12)	(0.13)	(0.13)	(0.08)	(0.09)
Middle	-0.15	-0.05	0.06	-0.10	-0.00	-0.08	-0.12	-0.26**	-0.13	-0.22***
	(0.09)	(0.07)	(0.09)	(0.08)	(0.15)	(0.07)	(0.11)	(0.09)	(0.13)	(0.06)
Height-for-age z-score at age 5 (vs. High)										
Low	0.01	0.01	0.02	0.10	0.03	-0.05	0.11	0.08	-0.06	0.09
	(0.14)	(0.11)	(0.12)	(0.11)	(0.07)	(0.06)	(0.10)	(0.11)	(0.12)	(0.07)
Middle	0.10	0.10	0.13**	0.11	0.01	0.07	0.21*	0.14***	-0.03	0.08
	(0.08)	(0.06)	(0.05)	(0.08)	(0.05)	(0.08)	(0.10)	(0.03)	(0.08)	(0.12)
Health prob. at age 5(vs. No)	-0.18	-0.38*	-0.15	-0.08	-0.07	-0.15	-0.01	-0.15	-0.19*	-0.04
	(0.16)	(0.17)	(0.12)	(0.11)	(0.14)	(0.16)	(0.09)	(0.19)	(0.10)	(0.11)
Father's highest education level (vs. Secondary and above)										
No education	0.03	-0.13	-0.10	-0.35***	-0.31**	-0.31***	-0.17	-0.13	-0.24**	-0.10
	(0.13)	(0.08)	(0.13)	(0.06)	(0.10)	(0.09)	(0.10)	(0.11)	(0.10)	(0.13)

Primary education	-0.06	0.06	-0.17**	-0.20***	-0.17**	-0.21**	-0.06	-0.08	-0.08	-0.00
	(0.09)	(0.08)	(0.05)	(0.05)	(0.05)	(0.07)	(0.11)	(0.10)	(0.08)	(0.09)
Caregiver's highest education level (vs. Secondary and above)										
No education	0.01	-0.13	-0.03	0.02	-0.05	0.07	0.09	-0.08	0.01	0.05
	(0.09)	(0.12)	(0.14)	(0.09)	(0.13)	(0.07)	(0.11)	(0.20)	(0.10)	(0.13)
Primary education	-0.02	-0.07	0.02	-0.01	-0.17	-0.04	0.07	-0.00	0.03	0.09
	(0.09)	(0.07)	(0.07)	(0.07)	(0.14)	(0.08)	(0.08)	(0.15)	(0.11)	(0.06)
Household size (> 6)	-0.12**	-0.09	-0.12	-0.16*	-0.16**	0.02	-0.13	0.00	0.01	-0.11
	(0.05)	(0.12)	(0.10)	(0.07)	(0.06)	(0.08)	(0.10)	(0.13)	(0.10)	(0.13)
First born	-0.05	0.04	0.02	-0.06	0.00	-0.03	0.12	0.06	0.12	0.15**
	(0.07)	(0.02)	(0.05)	(0.08)	(0.10)	(0.13)	(0.08)	(0.10)	(0.07)	(0.05)
Same language home/school	-0.04	-0.09	0.09	-0.01	0.09	0.07	0.04	-0.02	-0.08	-0.06
	(0.06)	(0.08)	(0.14)	(0.10)	(0.07)	(0.09)	(0.09)	(0.09)	(0.06)	(0.12)
Wealth quintile (vs. Quintile 5)										
Quintile 1	-0.20	-0.37**	-0.37**	-0.20	-0.40***	-0.43**	-0.16	-0.15	-0.16*	-0.20
	(0.12)	(0.14)	(0.11)	(0.12)	(0.09)	(0.17)	(0.12)	(0.14)	(0.08)	(0.15)
Quintile 2	-0.38**	-0.40**	-0.37**	-0.23	-0.33**	-0.31	-0.21	-0.09	-0.12	-0.19
	(0.12)	(0.13)	(0.13)	(0.14)	(0.13)	(0.20)	(0.13)	(0.20)	(0.10)	(0.13)
Quintile 3	-0.09	-0.27**	-0.10	-0.07	-0.26**	-0.24	-0.03	-0.07	-0.10	-0.12
	(0.11)	(0.09)	(0.12)	(0.11)	(0.11)	(0.19)	(0.11)	(0.13)	(0.11)	(0.16)
Quintile 4	0.12	-0.21	0.01	0.04	-0.27	-0.12	-0.02	-0.02	-0.05	0.00
	(0.11)	(0.13)	(0.08)	(0.09)	(0.15)	(0.12)	(0.07)	(0.16)	(0.14)	(0.15)
Region (vs. Addis Ababa)										
Living in Tigray	-0.27	-0.60***	-1.00***	0.59***	-0.39**	-1.01***	0.35	0.06	-0.38	-0.10
	(0.20)	(0.15)	(0.16)	(0.13)	(0.16)	(0.20)	(0.29)	(0.14)	(0.21)	(0.30)
Living in Amhara	0.05	0.27*	0.03	0.19	-0.03	-0.44***	0.16	0.27	0.38**	-0.24*
	(0.15)	(0.14)	(0.09)	(0.16)	(0.16)	(0.13)	(0.18)	(0.14)	(0.13)	(0.13)
Living in Oromia	-0.65***	-0.23*	-0.33*	-0.56***	-0.33*	-0.79***	-0.49**	-0.90***	-0.76***	-0.54***
	(0.15)	(0.10)	(0.16)	(0.14)	(0.15)	(0.10)	(0.15)	(0.13)	(0.09)	(0.10)
Living in SNNP	0.14	-0.34***	-0.27***	0.10	-0.52***	-0.77***	-0.18	-0.37***	-0.52***	-0.63***
	(0.09)	(0.10)	(0.05)	(0.09)	(0.14)	(0.06)	(0.12)	(0.11)	(0.07)	(0.12)
Child's ethnicity (vs. SNNP)										
Ethnicity 1: Others	0.28	0.40**	0.19***	0.22	0.24	0.14	0.14	0.44***	0.17	0.15
	(0.18)	(0.15)	(0.04)	(0.15)	(0.17)	(0.08)	(0.16)	(0.08)	(0.15)	(0.11)
Ethnicity 3: Oromo	0.34*	0.21	0.04	0.11	-0.12	-0.05	0.10	0.32*	0.11	0.13
	(0.17)	(0.17)	(0.06)	(0.09)	(0.11)	(0.09)	(0.18)	(0.16)	(0.08)	(0.23)
Ethnicity 4: Tigrian	0.19	0.40**	0.22**	0.16	0.12	0.34	0.32	0.24*	0.16	0.10
	(0.27)	(0.17)	(0.07)	(0.18)	(0.14)	(0.19)	(0.30)	(0.12)	(0.19)	(0.38)
Ethnicity 5: Amhara	0.34	0.32*	0.10**	0.14	-0.05	0.05	0.15	0.22**	0.21*	0.23*
	(0.19)	(0.17)	(0.04)	(0.15)	(0.10)	(0.16)	(0.18)	(0.08)	(0.11)	(0.11)

Constant	-2.13***	-0.96	0.79	-2.01***	0.78	0.99*	-1.68***	-1.80**	-1.56**	-0.81
	(0.47)	(0.61)	(0.51)	(0.49)	(0.57)	(0.52)	(0.48)	(0.62)	(0.56)	(0.47)
Observations	652	632	652	646	622	652	650	610	652	602
R-squared	0.39	0.35	0.34	0.35	0.20	0.30	0.22	0.24	0.34	0.34

Note: (1) Preschool coefficients and ATTs are based on the standardised score (z-score) of each test; (2) robust standard errors, clustered at community level, in parentheses; (3) Ethnicity in SNNP includes Hadiva, Sidama, Wolayta; and others includes Agew, Gurage, Kambata.
*** p<0.01, ** p<0.05, *p<0.1
Source: Young Lives Dataset Round 2 to Round 5, Young Lives

Table E. 3. Young Lives: Relation between Preschool and Educational Attainment, Full OLS/ Logit Model (Pooled/Urban sample)

Variables	POOLED SAMPLE				URBAN SAMPLE			
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
	Highest grade achieved		On-time progression (Odds ratio)		Highest grade achieved		On-time progression (Odds ratio)	
Round (Age)	Round 3 (age 8)	Round 4 (age 12)	Round 5 (age 15)	Round 3 (age 8)	Round 4 (age 12)	Round 5 (age 15)	Round 3 (age 8)	Round 4 (age 12)
Preschool attendance	0.62***	0.56***	2.04**	3.39***	0.40*	0.37	1.82	2.64***
	(0.15)	(0.18)	(0.60)	(0.66)	(0.18)	(0.24)	(0.72)	(0.68)
Private Spending on education (vs. high)								
Low	-0.06	-0.16	1.15	1.01	-0.39***	-0.31***	1.13	0.75**
	(0.09)	(0.10)	(0.19)	(0.14)	(0.07)	(0.05)	(0.24)	(0.10)
Middle	0.09	0.03	0.99	1.05	0.21**	0.26*	0.97	1.45
	(0.06)	(0.08)	(0.14)	(0.13)	(0.09)	(0.12)	(0.14)	(0.33)
Parental aspiration	0.19	0.28*	1.48**	1.55**	-0.04	-0.13	1.46*	1.25
	(0.13)	(0.15)	(0.25)	(0.28)	(0.12)	(0.14)	(0.30)	(0.38)
Age 14 (vs. Age 15)	-0.43***	-0.26***	1.00	0.78*	-0.16	-0.17	1.42**	1.09
	(0.07)	(0.07)	(0.15)	(0.10)	(0.16)	(0.16)	(0.24)	(0.38)
Female (vs. Male)	0.23**	0.44***	1.34	1.65***	-0.30**	-0.35*	0.91	0.45**
	(0.09)	(0.12)	(0.27)	(0.28)	(0.12)	(0.16)	(0.26)	(0.15)
PPVT at age 5 (vs. High)								
Middle	-0.03	-0.04	1.26*	0.85	-0.11	-0.15	0.97	0.78
	(0.07)	(0.10)	(0.17)	(0.13)	(0.14)	(0.13)	(0.26)	(0.18)
High	0.10	0.18	1.17	1.12	0.26***	0.20**	1.50***	1.30
	(0.11)	(0.11)	(0.20)	(0.25)	(0.06)	(0.08)	(0.22)	(0.31)
CDA-Q(Math) at age 5 (vs. High)								
Low	-0.15	-0.20	1.21	0.87	0.16*	0.14	1.52***	1.06
	(0.16)	(0.17)	(0.29)	(0.23)	(0.08)	(0.14)	(0.22)	(0.33)
Middle	-0.13	-0.20**	1.11	0.80*	-0.17	-0.46*	1.23	0.68
	(0.09)	(0.09)	(0.20)	(0.10)	(0.15)	(0.20)	(0.45)	(0.38)
Height-for-age z-score at age 5 (vs. High)								
Low	-0.49***	-0.44***	0.58***	0.56***	-0.31*	-0.32*	1.53	0.78
	(0.08)	(0.10)	(0.09)	(0.10)	(0.14)	(0.15)	(0.51)	(0.20)
Middle	-0.14*	-0.19	0.83	0.73	-0.13	-0.12	0.94	0.75
	(0.08)	(0.11)	(0.11)	(0.16)	(0.13)	(0.17)	(0.23)	(0.17)
Health prob. at age 5(vs. No)	-0.17	-0.26	0.89	0.75	0.13	-0.04	0.75	0.98
	(0.12)	(0.17)	(0.17)	(0.18)	(0.17)	(0.14)	(0.32)	(0.37)

Father's highest education level (vs. Secondary and above)								
No education	-0.26**	-0.40**	1.18	0.64**	0.08	-0.05	1.08	0.87
	(0.12)	(0.15)	(0.28)	(0.14)	(0.06)	(0.10)	(0.33)	(0.26)
Primary education	-0.25**	-0.30**	1.01	0.72	-0.13	-0.14	1.09	0.99
	(0.11)	(0.14)	(0.22)	(0.15)	(0.11)	(0.13)	(0.18)	(0.21)
Caregiver's highest education level (vs. Secondary and above)								
No education	0.08	-0.08	0.70	0.88	0.09	0.09	0.94	0.93
	(0.13)	(0.13)	(0.22)	(0.27)	(0.10)	(0.14)	(0.22)	(0.25)
Primary education	0.08	-0.03	1.12	0.97	0.02	0.02	0.88	1.34
	(0.08)	(0.11)	(0.30)	(0.25)	(0.14)	(0.17)	(0.20)	(0.40)
Household size (> 6)	-0.19**	-0.14	0.83	0.85	-0.03	0.12	1.35	1.50
	(0.07)	(0.08)	(0.10)	(0.13)	(0.14)	(0.21)	(0.40)	(0.45)
First born	0.01	0.22*	0.93	0.98	-0.36**	-0.58***	1.30	0.43**
	(0.10)	(0.12)	(0.15)	(0.17)	(0.15)	(0.12)	(0.38)	(0.16)
Same language home/school	-0.07	-0.08	0.71	1.15	-0.22	-0.35	1.20	0.55
	(0.13)	(0.14)	(0.15)	(0.25)	(0.20)	(0.25)	(0.24)	(0.27)
Wealth quintile (vs. Quintile 4)								
Quintile 1	-0.49***	-0.75***	0.33***	0.31***	-0.15	-0.23	1.07	0.70
	(0.17)	(0.19)	(0.08)	(0.08)	(0.19)	(0.17)	(0.25)	(0.29)
Quintile 2	-0.15	-0.38**	0.66**	0.62*	0.00	-0.03	0.82	0.86
	(0.14)	(0.16)	(0.12)	(0.17)	(0.13)	(0.16)	(0.19)	(0.14)
Quintile 3	-0.17*	-0.18*	0.67**	0.71	-0.02	-0.02	0.95	1.26
	(0.09)	(0.10)	(0.12)	(0.15)	(0.18)	(0.22)	(0.26)	(0.38)
Quintile 5	0.12	0.17	0.70**	1.22	-0.00	0.01	1.04	1.30
	(0.08)	(0.11)	(0.12)	(0.32)	(0.12)	(0.15)	(0.19)	(0.29)
Region (vs. Addis Ababa)								
Living in Tigray	1.49***	1.54***	3.44***	7.48***	1.07**	1.31***	3.01***	5.00***
	(0.38)	(0.42)	(1.05)	(4.67)	(0.33)	(0.38)	(1.03)	(2.59)
Living in Amhara	0.27	0.44	2.06*	2.50**	0.93**	1.17***	1.70	6.13***
	(0.32)	(0.32)	(0.85)	(0.95)	(0.30)	(0.30)	(0.65)	(2.69)
Living in Oromia	-0.25	-0.11	1.18	0.78	-0.36*	-0.25	1.01	0.63
	(0.20)	(0.22)	(0.32)	(0.21)	(0.17)	(0.19)	(0.27)	(0.24)
Living in SNNP	-0.98***	-0.77***	0.86	0.44***	-0.68***	-0.63***	1.03	0.42***
	(0.19)	(0.20)	(0.30)	(0.09)	(0.16)	(0.15)	(0.26)	(0.11)
Child's ethnicity (vs. SNNP)								
Ethnicity 1: Others	-0.17	0.07	0.94	1.82*	0.33	0.44*	1.61	2.21
	(0.18)	(0.17)	(0.30)	(0.60)	(0.20)	(0.23)	(0.77)	(1.25)

Ethnicity 3: Oromo	-0.05	0.10	1.52*	1.71	0.37	0.30	2.09***	1.49
	(0.17)	(0.16)	(0.36)	(0.56)	(0.21)	(0.23)	(0.55)	(0.67)
Ethnicity 4: Tigrian	-0.05	0.29	2.41***	2.52**	0.19	0.26	3.12***	1.96
	(0.33)	(0.35)	(0.71)	(0.96)	(0.38)	(0.39)	(1.35)	(0.87)
Ethnicity 5: Amhara	0.08	0.24	1.47	1.59*	0.26	0.36	1.80*	1.40
	(0.15)	(0.15)	(0.35)	(0.40)	(0.24)	(0.27)	(0.60)	(0.56)
urban	0.42	0.41	1.56**	1.45	-	-	-	-
	(0.25)	(0.27)	(0.31)	(0.53)	-	-	-	-
Constant	-2.13***	-0.96	0.79	-2.01***	0.78	0.99*	-1.68***	-1.80**
	(0.47)	(0.61)	(0.51)	(0.49)	(0.57)	(0.52)	(0.48)	(0.62)
Observations	1,402	1,447	1,447	1,447	649	652	652	652
R-squared	0.43	0.38	-	-	0.38	0.37	-	-

Note: (1) Model 1, 2, 5, 6 use OLS regression, and standard errors, clustered at community level, in parentheses; (2) Model 3, 4, 7, 8 use Logit regression, and standard errors, clustered at community level, in parentheses; (3) Ethnicity in SNNP includes Hadiva, Sidama, Wolayta; and others includes Agew, Gurage, Kambata..
*** p<0.01, ** p<0.05, *p<0.1
Source: Young Lives Dataset Round 2 to Round 5, Young Lives

Table E. 4. Young Lives: Results of SEM with Observable School Characteristics

	(1) PPVT- Round 4	(2) PPVT- Round 5
Age	(age 12)	(age 15)
Path coefficients **(S.E)**		
Preschool attendance → School Environments → PPVT outcome		
Total effects	1.04***	1.02***
(SE)	(0.09)	(0.09)
Direct effects	0.92***	0.94***
(SE)	(0.08)	(0.09)
Indirect effects	0.12***	0.08***
(SE)	(0.03)	(0.03)
% of total effect mediated	11.5	7.8
% of total effect unmediated	88.5	92.2
Ratio of indirect effect to direct effect (R*m*)	0.13	0.09
Model Fit		
Chi-Square (*DF*)	61.01(6)	62.87(6)
CFI	0.80	0.75
RMSEA	0.14	0.14
SRMR	0.07	0.07
R-Square	0.24	0.23

Note: (1) The resulting structural coefficients (standardised regression coefficients) describe the direct and indirect effects. (2) Standard errors are in parentheses; (3) DF: Degree of Freedom; (4) CFI: Comparative Fit Index; (5) RMSEA: Root Mean Square Error of Approximation; (6) SRMR: Standardised Root Mean Square Residual.
*** p<0.01, ** p<0.05, *p<0.1
Source: Young Lives Dataset Round 2 to Round 5, Young Lives School Survey Dataset 2012-2013, Young Lives