The Relationship Between Literacy Outcomes and Social-Contextual Variables for Students in Low– and Middle–Income Countries

Candace Debnam

A Research & Development Project

Submitted for the MSc Educational Assessment 2022

DEPOSIT AND CONSULTATION OF THESIS

One copy of your dissertation will be deposited in the Department of Education Library via Oxford Research Archives where it is intended to be available for consultation by all Library users. In order to facilitate this, the following form should be inserted in the library copy of the dissertation.

Note that some graphs/tables may be removed in order to comply with copyright restrictions.

Surname	Debnam
First Name	Candace
Faculty Board	Education
Title of Dissertation	The Relationship Between Literacy Outcomes and Social-Contextual Variables for Students in Low- and Middle-Income Countries

Declaration by the candidate as author of the dissertation

1. I understand that I am the owner of this dissertation and that the copyright rests with me unless I specifically transfer it to another person.

2. I understand that the Department requires that I shall deposit a copy of my dissertation in the Department of Education Library via Oxford Research Archives where it shall be available for consultation, and that reproductions of it may be made for other Libraries so that it can be available to those who to consult it elsewhere. I understand that the Library, before allowing my dissertation to be consulted either in the original or in reproduced form, will require each person wishing to consult it to sign a declaration that he or she recognises that the copyright of this thesis belongs to me. I permit limited copying of my dissertation by individuals (no more than 5% or one chapter) for personal research use. No quotation from it and no information derived from it may be published without my prior written consent and I undertake to supply a current address to the Library so this consent can be sought.

3. I agree that my dissertation shall be available for consultation in accordance with paragraph 2 above.

The Relationship Between Literacy Outcomes and Social-Contextual Variables for Students in Low- and Middle-Income Countries

ABSTRACT: This dissertation explores the relationship between social-contextual variables and literacy outcomes in Sierra Leone. To ensure the quality of the review, the assessment tool utilized to measure literacy outcomes, in this case, the Early Grade Reading Assessment (EGRA), will be evaluated for appropriateness, using an examination of descriptive statistics, Rasch methodology, and correlations with social-contextual variables. While the assessment may be of acceptable quality, student reading levels are low. However, despite these low scores, students' achievement was potentially positively impacted by reading with parents and the language spoken at home and at school.

Table of Contents

List of Abbreviations

APFL: All Pikin for Learn

CTA: community teacher association

DIBELS: Dynamic Indicators of Basic Early Literacy Skills

DIF: differential item functioning

EGRA: Early Grade Reading Assessment

GoSL: Government of Sierra Leone

ICC: item characteristic curve

IRT: item response theory

LMICs: low- and middle-income countries

MSG: mothers' support group

PSI: Person Separation Index

SMC: school management committee

UNESCO: United Nations Educational, Scientific and Cultural Organization

USAID: United States Agency for International Development

List of Tables

List of Figures

The Relationship Between Literacy Outcomes and Social-Contextual Variables for Students in Low- and Middle-Income Countries

1 Introduction

Education gives people the means to lead fulfilling lives, flourish personally, and contribute to their communities. In addition, education makes it more likely that a person can successfully access healthcare, find work that pays a living wage, and live in a safe, unpolluted environment—all factors that affect well-being (United Nations Educational, Scientific and Cultural Organization [UNESCO], 2019a). Few can deny education as a vital social determinant when many factors contribute to a person's well-being (Zajacova & Lawrence, 2018). Therefore, understanding the social-contextual variables that impact student educational outcomes is vital to determining how to best structure future education programming and decisions. This dissertation explores the relationship between social-contextual variables and literacy outcomes in Sierra Leone. To ensure the quality of the review, the assessment tool utilized to measure literacy outcomes, in this case, the Early Grade Reading Assessment (EGRA), will be evaluated for appropriateness.

1.1 Literacy Skills and Sustainable Development

Reading is a critical foundational skill for academic attainment and careers in the current knowledge economy. Because of this, the development of reading competencies has significant repercussions for individual children. This importance is highlighted by the experiences of students who enter the upper primary grades with poor reading skills. These students often struggle throughout their schooling (Francis et al., 1996). Additionally, they exhibit a high probability of educational failure, dropout, joblessness, overall low income, unlawful action, and

poor health (Christenson & Thurlow, 2004; Hernandez, 2011). Outside of growth and development for a person, learning to read also has economic effects at the community level. For example, Hanusheck and Woessmann found that over 70% of the variation in economic growth at the country level can be attributed to a population's performance on assessments, particularly in reading (2012).

Regrettably, many students, specifically those in low- and middle-income countries (LMICs), do not develop foundational reading skills even after several years of instruction (UNESCO, 2019b). For example, according to the United Nations 2018 report, 9 of 10 children (i.e., 202 million children) in sub-Saharan Africa are unable to demonstrate basic reading skills. An equally alarming situation can be seen in south and central Asia, where over 80% of children (or 241 million children) cannot meet minimum proficiency standards for literacy skills (UNESCO, 2018). In addition, the World Bank estimates that over 80% of children in developing countries cannot read and understand a simple story (World Bank, 2019). In Sierra Leone, achieving quality reading instruction has been particularly slow. This can be seen in the data presented in the Multiple Indicator Cluster Survey Round 6 of Statistics Sierra Leone (2018), which revealed that nearly 85% of Sierra Leone students could not read a simple story. In response to these unsettling findings, stakeholders in Sierra Leone and other LMICs have, over the last decade, paid unparalleled attention to reading instruction in primary grades. This can be seen through the various international initiatives such as the Sustainable Development Goals, the Education for All initiative by United Nations agencies, and other actions by donors and nongovernmental organizations.

The importance of literacy as a critical social driver was solidified in the ambitious Millennium Development Goals in 2015. Based on the 2030 vision, United Nations Sustainable

Development Goal 4 recommends that all nations offer comprehensive quality education and advocate for long-term learning for every child (Torrente et al., 2019). As a result, numerous countries have introduced school interventions to build students' reading skills, particularly in LMICs. These interventions have focused on improving early-grade reading outcomes by grounded new interventions in the "simple view of reading."

1.2 The Simple View of Reading

Learning to read is a multifaceted process involving synchronizing a wide range of skills. The simple view of reading theorizes that reading proficiency can be understood by examining the interaction between fluency and comprehension—and interventions focused on improving student outcomes should be focused on building these two skills. This distillation of reading was a break from the prior, more complex framings of reading. Before 1986, the understanding of reading as a complicated activity had been a cornerstone aspect of the understanding and development of literacy pedagogical approaches. In the early 1900s, Huey (1908/1968) posited that analyzing reading could best be described as "very many of the most intricate workings of the human mind" (p. 6). Forty years later, Gates (1949) stated a parallel view, declaring that reading is "a complex organization of patterns of higher mental processes . . . [that] . . . can and should embrace all types of thinking, evaluating, judging, imagining reasoning, and problem-solving" (p. 3). The creators of a report authorized by the National Academy of Education took a similar stance when they compared reading to "the performance of a symphony orchestra" (Anderson et al., 1985, p. 7).

When Gough and Tunmer (1986) conceived the simple view of reading in a short paper nearly 40 years ago, it is unlikely that the researchers had a grasp of the influence that their simple but astute framing would have on the subject of reading. Their original study followed a

cohort of over 250 students from kindergarten through early primary school. Students were assessed annually from 1978-1985 using several assessments to track individual students' development of reading skills. The influence of the simple view of reading grew slowly over time, but its impact grew exponentially in the early 2000s. Bartlett et al. (2015) traced the explosion of focus on the simple view of reading to an influential report released in 2000 by the U.S. National Reading Panel called *Teaching Children to Read*. This controversial yet influential report streamlined complicated questions about reading acquisition and language development. The report developed the idea of the five pillars of reading:

1. phonemic awareness, which can be understood as the ability to distinguish the individual sounds in spoken words;
2. phonics, which can be understood as grasping the connection between letters (graphemes) and sounds (phonemes);
3. fluency, which can be understood as the ability to read text both quickly and accurately;
4. vocabulary, which can be understood as the body of words known to a person;
5. comprehension, which can be understood as the ability to recognize and convey meaning from what is read.

While the importance of these skills are indisputable, some academics have criticized the way scientific evidence has been "reviewed, distorted, and misrepresented" throughout the National Reading Program (NRP) report. This critique is particularly important because of the numerous policies and programming decisions drawn from the report (Allington, 1999, p. 22). For example, the "five core components" of literacy do not include some components that stakeholders feel represent crucial aspects of reading, such as print and lexical knowledge

(Cunningham et al., 2002). Alternately, other stakeholders felt the report simplified complex disputes about fundamental terms, notably phonics (e.g., Garan, 2002), fluency (e.g., Krashen, 2002), and the connection between phonics and comprehension (e.g., Coles, 2000), allowing for a more unified and effective policy framework to emerge. For example, the Reading First guidelines, part of the U.S. reform No Child Left Behind, were deeply grounded in this five-point model. This was principally reflected in the increased focus on fluency. These guidelines also aligned with the growing international focus on improving literacy globally and began to inform international literacy intervention design and assessment, particularly in interventions funded by United States Agency for International Development (USAID, 2017).

While the simple view presents reading as this specific two-way interaction between fluency and comprehension, it maintains an understanding of reading as a complex and nuanced process. Comprehension and decoding are universally viewed as complicated processes. The simple view argues that these intricacies can be divided into two equally important parts (Hoover & Gough, 1990). Decoding, sometimes called word recognition, is defined "as efficient word recognition" (Hoover & Gough, 1990, p. 72). This definition goes further than the conventional description of decoding, which typically holds that decoding is the ability to sound out words using phonics. The definition of decoding is often expanded to include the "quick and accurate reading of both familiar and unfamiliar words in both lists and connected text" (Gough & Tunmer, 1986, p. 84).

Comprehension is often viewed as more complicated as it includes the tenant of language comprehension as an underpinning requirement of reading comprehension. Language comprehension, the foundational concept, is often called by other names, such as linguistic or listening comprehension. All of these terms relate to a student's ability to derive meaning from

spoken words. This type of language comprehension encompasses "receptive vocabulary, grammatical understanding, and discourse comprehension" (Catts, Adlof, & Weismer, 2006). Reading comprehension and language comprehension differ because reading comprehension relies on written rather than oral language and centers on a student's ability to distinguish the words and derive meaning (Hoover & Gough, 1990). Ultimately, this means that language comprehension develops into reading comprehension when the meaning of a word is derived from print rather than oral language. Because of this, it is possible, and in some situations likely, that students would be able to demonstrate strong language comprehension and still demonstrate low reading skills if there are challenges with decoding.

Additionally, Kamhi (2007) articulately defines the distinction between reading comprehension and decoding. According to Kamhi, decoding is "a teachable skill" compared to comprehension, which "is not a skill and is not easily taught" (p.54). Kamhi expounds that word recognition is a teachable skill because it "involves a narrow scope of knowledge (e.g., letters, sounds, words) and processes (decoding) that, once acquired, will lead to fast, accurate word recognition" (p.54). Furthermore, Kamhi states that comprehension "is not a skill. It is a complex of higher-level mental processes that include thinking, reasoning, imagining, and interpreting" (p.55).

Since the beginning of the simple view of reading, hundreds of studies have been built using this model as a framework to direct their research and understand their results. Many studies have investigated the model's central idea that reading results from the interaction between comprehension and decoding. This work has established that a great deal of the variation in reading comprehension can be explained by individual variations in decoding and language comprehension (Catts & Kamhi, 2005). This is the case for both alphabetic

orthographies, such as English, Greek (Protopapas et al., 2012), and Italian (Tobia & Bonifacci, 2015), as well as non-alphabetic writing systems like Chinese (Ho et al., 2007). Because of these similarities across orthographies, a subset of assessment designers with support from several bilateral and multi-lateral donors began exploring the possibility of developing an assessment framework based on the simple view of reading that could easily be used in developing country contexts.

1.3 An Introduction to the EGRA

1.3.1 Development and Theoretical Underpinning of the EGRA

As nations began to launch new education interventions based on the simple view of reading, the international community began developing an assessment that could be used to effectively understand the impact of these new interventions on students' progress toward literacy. In 2006, USAID conducted a desk review of early-grade reading that embraced the five-pillar model of reading. The resulting report pulled heavily from the National Research Council's *Preventing Reading Difficulties in Young Children* and the National Reading Panel's *Teaching Children to Read.* The subsequent recommendations endorsed the development of a new assessment for use in LMICs that relied on markers like "words correct per minute." However, the recommendations also underscored the significance of pre-reading skills, learning materials, contact with teachers or other instructors, as well as mother-tongue literacy and oral language development (Bartlett et al., 2015). In 2006, USAID supported the development of this recommended tool for assessing early grade reading. An international expert panel took the Dynamic Indicators of Basic Early Literacy Skills (DIBELS) subtests, modified them to some extent, and piloted them in the context in which they were intended to be used. The result of this work was the EGRA (Gove & Cvelich, 2011). This final tool was designed to be used in LMICs

to measure students' skills at both pre-reading and reading —and was intended to serve several purposes, ranging from serving as a baseline of early reading acquisition, to serving as a large-scale education systems analysis, to a formative tool used to inform instruction.

As noted, the test developers drew heavily on the DIBELS assessment, which serves as the basis for EGRA. DIBELS is a classroom assessment developed in the United States (Samuels, 2007). While widely used, DIBELS has been criticized for making claims not based on evidence, misrepresenting the skills required to read, and testing only a portion of those skills. Additionally, other critics have disparaged DIBELS for focusing too much on speed over accuracy. Another standard critique is that the nature of an individually enumerated assessment can mean it is difficult to administer consistently. Finally, concerns over DIBELS benefiting financially from the inappropriate promotion of tests as part of the federal Reading First program in the wake of No Child Left Behind have also been noted as a concern (Kamii & Manning, 2005; Samuels, 2007; Shelton et al., 2009). Many of these same criticisms have been leveled against the EGRA.

Despite these concerns, since its release, over 65 countries have implemented the DIBELS-based EGRA, resulting in its adaptation into more than 100 languages (Dubeck & Gove, 2015). While the EGRA's developers intended the assessment to serve many purposes, even they recognize that it is not a perfect assessment. However, within the framework of the Education 2030 agenda and the Sustainable Development Goals, the foundations and influence of the EGRA are exceptionally important to consider. According to Bartlett et al. (2015), the EGRA currently exerts "matchless influence on assessment policies" (p. 36). This impact can be seen in how it has been used by prominent donors, NGOs, and academics (Halliday et al., 2012). One specific example of this influence can be seen in the *Early Reading: Igniting Education for*

All report (Gove & Cvelich, 2011), which reviewed EGRA results from multiple regions. These results served as the basis for extensive program content funded by large-scale, international actors.

1.3.2 Composition, Strengths, and Weaknesses of the EGRA

Logistically, the EGRA is an oral assessment individually administered to students in the early grades to measure the most foundational skills for literacy achievement. The assessment concentrates on what it labels the "three early stages of reading acquisition":

Table 1

Early Grade Reading Assessment Components (adapted from RTI)

Stage	Test components [a]
Stage 0: Emergent literacy Birth to Grade 1	• Concepts about print • Phonemic awareness • Listening comprehension
Stage 1: Decoding Beginning Grade 1	• Letter naming • Letter sounds • Syllable naming • Nonsense word reading • Familiar word reading
Stage 2: Confirmation and fluency End of Grade 1 to end of Grade 3	• Paragraph reading (oral reading fluency) with comprehension • Dictation

[a] Not all components are tested in all languages.

The most significant difference in skills are often seen in emergent (Stage 0) and decoding (Stage 1) skills. In contrast, confirmation and fluency (Stage 3) scores are frequently limited due to low student performance. Due to this, decoding, phonics, and phonemic awareness often receive a stronger focus in EGRA assessment practices when compared with comprehension. The most used task is the reading comprehension task which is individually assessed utilizing a subtask consisting of questions based on the oral reading fluency subtask.

Typically, the student will be given 1 minute to read a brief story aloud as the oral reading fluency subtask. After completing the oral reading fluency subtask, the enumerator will ask the student up to five reading comprehension questions. The student is only asked questions about the story when they stopped reading. This is the reading comprehension subtask. With only five items, students can achieve a score of 0%, 20%, 40%, 60%, 80% or 100% correct. While this method has the benefit of being straightforward and fast to administer, it has been criticized for generating high numbers of zero scores, which can be tricky to interpret (Hoffman, 2012). High levels of zero scores mean there are many students for whom limited or no information is available, creating a reduced understanding of where they are on their journey to reading acquisition.

The link between oral reading fluency and reading comprehension measures has also been questioned. For example, some scholars note that it is possible for a student to read slowly but also demonstrate a high level of accuracy and comprehension if they are given extended time to complete the subtasks (Dowd et al., 2020). If true, then the time limits employed during an EGRA administration would create an artificially deflated number of readers who may have been able to demonstrate comprehension if given additional time to complete the reading.

Wagner et al. (2012) pointed out that there is widespread consensus among test developers regarding what makes a "good" assessment. Despite this, some key parameters are still contested, especially in LMIC countries where the growth in assessments is quickly escalating. These contested parameters are often distilled into: "who gets tested, what gets tested, how tests are conducted, and why the test is conducted"(p. 35). These concerns are deeply interconnected. For example, "if minority language speakers or refugees are included in the

assessment population (the "who"), it is likely to have implications for the other issues, such as the type of content (the "what") included in the assessment" (p.36).

Wagner et al. (2012) and Gove and Wetterberg (2011) pointed out that EGRA-type assessments offer the following possibilities: The EGRA focuses on early-grade learning and interventions, unlike large-scale educational evaluations, which generally do not target children until fourth grade and often much later. At these later points, students may be much farther behind in reading development and there are fewer opportunities for early interventions that are crucial in terms of social justice and equity. Additionally, given the nature of the assessment, the minimum sample is often smaller since the EGRA is designed to serve as a monitoring tool, rather than as a high-stakes assessment requiring a nationally representative sample (Gove, 2017). Perhaps most importantly, the time between design and reporting can be considerably less than in other international assessments. Currently, the cost per student for administering the EGRA appears in line with other large-scale educational assessments. Though, assessment advocates note that costs may decrease as the tool become more familiar and staff are better trained (Gove, 2017).

Despite these positive points, there are also limitations to the EGRA that must be addressed. One is the overall conception of reading and language development underpinning the assessment design. The stage model of reading underlying the EGRA, described in detail above, has been extensively criticized in the literature is the idea that literacy is acquired in stages. This is not a universally embraced stance. For example, many respected reading experts assert that comprehension and fluency must be taught concurrently with decoding skills (Bartlett et al., 2015).

There is also the critique that the EGRA is biased toward phonics and that the EGRA has isolated reading components. By isolating these components, the assessment does not adequately acknowledge that these skills develop together. This perceived bias toward phonics is exemplified in the contentious task where children read "nonsense words." This subtask is often highlighted as an instance where the assessment structure sets phonics against comprehension (Bartlett et al., 2015). The EGRA is also criticized for framing reading as an isolated sphere within language development, rather than as an integrated component. For example, the EGRA glosses over the critical and developmentally important relationships between "print exposure, reading, and oral language development, as well as the links between reading and writing" (p,73). (Bartlett et al., 2015; Wagner et al., 2012).

An additional critique of the EGRA is its use of timed tasks. As noted, the assessment requires a student to read a passage aloud for 1 minute and then answer questions centered on that passage. A stopwatch is often used to track this. While stopwatches can assure uniformity in the administration of the assessment, they can also give an unintended signal to students, teachers, and administrators that speed is a critical benchmark of good reading (Wagner et al., 2012). As noted, this timing also means slower readers may not be able to demonstrate their comprehension fully. The underlying assumption is that fluency serves as evidence of comprehension (Abadzi & Centanni, 2020). However, this critique can be problematic since multilingual children, particularly if they are learning in a transparent language, have been found to decode quickly but lack comprehension (Cao et al., 2019; Dowd et al., 2020; Moore et al., 2017). Indeed, a recent study of words correct per minute and comprehension scores in two European languages (English and Dutch) and two African languages (Sabaot and Pokomo)

showed similar comprehension scores with varied rates of words correct per minute (Graham & van Ginkel, 2014).

Wagner et al. (2012) argued that assessments like the EGRA have distinct strengths since they can be adapted to local contexts and languages, unlike large-scale educational assessments like the Program for International Student Assessment. However, Bartlett et al. (2015) and Trudell and Schroeder (2007) note that this potential strength is often unexploited because meaningful adaptation is not consistently implemented. This lack of meaningful adaptation turns a potential strength of the EGRA into a weakness.

Additionally, distinct languages require unique reading strategies, and EGRA-based approaches may not always be appropriate for a language. For example, Bartlett et al. (2015) asked if it is relevant to use the stage model of reading, which is grounded in Anglo-centric reviews of the empirical evidence regarding how monolingual English-speaking children learn to read, to evaluate the students' progress toward developing reading skills in other languages and scripts. Trudell and Schroeder (2007) contended that taking local contexts and languages seriously is vital to fully exploiting EGRAs strengths. If meaningful adaptation is not undertaken, then the utility of the EGRA is vastly diminished.

Thus, learning to read varies by language, script, histories of pedagogies, student motivations (Baker & Wigfield, 1999), and students' and teachers' beliefs about reading. There may be universal cognitive processes involved in literacy learning, but they are expressed through economic, social, and political contexts. They are shaped by reading environments, teaching and learning materials, and pedagogies. Sociocultural beings express them—for example, young children socialized into a language in ways that do or do not emphasize rhyming words and initial sounds or build vocabulary associated with schooling. These dynamic

interrelationships necessarily shape the manifestation of the cognitive processes involved in literacy learning. There are good reasons to remain skeptical about universalist reading pedagogies or assessments.

Another potential weakness is the local capacity required to administer an EGRA effectively. For example, high-level skills are required to construct, plan, and manage assessments. Globally, these skills are in short supply, particularly in LMICs. Because a wide variety of assessments are available today, officials are often faced with tricky decisions regarding how to invest in choosing, developing, and adapting assessments to national and local contexts and languages. Guaranteeing that policymakers and key stakeholders have correct information on the real costs of assessments is a crucial step in identifying suitable tools to inform and influence programs aimed at improving educational outcomes (Wagner et al., 2011).

1.3.3 EGRA's Relationship to Other International Assessments

In contrast to many international assessments, the EGRA is currently not viewed as comparable in the traditional psychometric sense across different administrations. This means cross-country comparisons are difficult, if not impossible, when using EGRA data. This lack of comparability is because a new test is often developed to be contextually specific for each administration and that version is not psychometrically equated to other EGRAs in a manner that would allow for comparability. So, each EGRA is, in theory, tailored and developed to align with the unique linguistic context and curriculum within any administration area. Again, this contrasts with other international assessments, such as PISA or TIMMS, which are often explicitly designed to be comparable across administration timepoints and countries.

Wagner et al. (2012) argued that the EGRA represents a hybrid type of assessment. First, Wagner et al. put the EGRA into context and distinguished between "four main types of

assessments: national, regional, international, and hybrid" (p.32). They pointed out that hybrid assessments have focused on the needs of low-income countries' assessment contexts in recent years. Hybrid assessments were initially conceptualized as "smaller, quicker, cheaper" reading assessment methods. The basic idea of a hybrid assessment is to see whether large-scale educational assessment methodologies could be transformed into hybrid methods that are just big enough to be informative, quicker at capturing, analyzing, and disseminating data, and cheaper in terms of personnel and other cost outlays (Wagner et al., 2011). With their particular intent, EGRA-like assessments often provide advantages not obtainable with other assessment types in terms of size, suitability, and cost-efficiency. However, this may not always be the case, given the adaptation and piloting necessary for each new EGRA administration.

While the EGRA has is strengths and weaknesses, it was developed to help to build a deeper understanding of the reading skills progression of children in developing countries. While building this understanding is crucial to allow governments and funders to design and fund reading interventions, other factors linked to literacy outcomes must be explored to understand students' learning context fully.

1.4 The Links Between Social-Contextual Factors and Literacy Outcomes

Many studies have shown the considerable influence of contextual factors on students' academic performance (e.g., Chen et al., 2021; Creemers & Kyriakides, 2010). These studies are often grounded in Walberg's educational productivity theory and Bronfenbrenner's ecological system theory which highlights that human learning can only be understood by considering the

impact of multiple factors. This study will examine the contextual factors that influence students' reading performance.

Bronfenbrenner's (1979) ecological system model is extensively used to investigate the impact of contextual factors on students (e.g., Eriksson et al., 2018; Leonard, 2011). The model underscores that a student's environment affects how the child develops (Bronfenbrenner, 1979). Likewise, Walberg's (1984) educational productivity model highlights the influence of contextual factors on individuals' learning. Both Bronfenbrenner's ecological system and Walberg's educational productivity recognize that students' learning cannot be grounded on a single factor and propose that learning can only be achieved when factors are linked (Bronfenbrenner, 1979; Walberg, 1984). Three main factors will be examined – 1) student age, grade level, and links to early learning, 2) student language, and 3) home environment.

1.4.1 Student Age, Grade Level, and Links to Early Learning

In August 2018, the Government of Sierra Leone (GoSL) launched a phased Free Quality School Education (FQSE) initiative that provides free admission and tuition to all children in government-approved schools. FQSE included the provision of pre-primary education. As of 2018, 21% of primary-age school children (6-11 years) and 29% of secondary school-age children (12-17) were out of school (UNESCO, 2020). Because of the sizeable out-of-school population and the relatively recent launch of universal primary education, there is a backlog of students progressing through the system leading to high numbers of overage students. While this structural constraint is likely to contribute to the overage population, teaching quality is often an additional contributing factor. Repeating years can cause students to become demotivated and drop out of school altogether, increasing the economic burden on families. Additionally, longitudinal studies point to additional benefits that arise from high-quality pre-primary

education, such as higher secondary school graduation rates, less retention in lower grades, academic achievement throughout schooling, and increased economic return on investment (Duncan et al., 2007; Elango, Garcia, Heckman, & Hojman, 2015; McCoy et al., 2017; Yoshikawa et al., 2013).

Considerable evidence from high-income countries (HICs) and growing evidence from low- and middle-income countries (LMICs) supports the immediate and long-term benefits of attending at least one year of high-quality pre-primary education (Earle, Milovantseva, & Heymann, 2018; McCoy et al., 2017). The benefits of high-quality pre-primary education are typically even more significant for learners from marginalized and vulnerable populations, who may not have the same opportunities as their peers to develop the skills, attitudes, and behaviors that prepare them for school (van Huizen & Plantenga, 2018; Yoshikawa et al., 2013). There is also evidence that pre-primary attendance supports specific school readiness skills, including language and literacy, numeracy, and social-emotional development (UNICEF, 2019; MacDonald & Murphy, 2019; Arapa et al., 2021). Long-term benefits include greater educational attainment, health, and wealth (Krafft, 2015). For example, school readiness programs were associated with lower rates of grade repetition in Nepal and lower rates of primary school drop-out in Cambodia (Nonoyama-Tarumi & Bredenberg, 2009). A large, longitudinal study in Chile indicated that children who attended at least one year of pre-school demonstrated significantly higher math and reading scores in fourth grade (Cortazar, 2015). A recent meta-analysis of the effects of literacy interventions in LMICs, a majority of which targeted literacy instruction, suggests that programs were most effective for emergent literacy skills (Kim, Lee, & Zuilkowski, 2020). Despite this evidence, most children in LMICs do not

have access to pre-primary education and, therefore, are not prepared to begin primary school (UNESCO, 2018).

There is also evidence that large populations of overage or underage students can be detrimental to overall student performance. Although grade retention is a standard measure to help students catch up, research suggests that young people do not generally benefit from repeating a year for academic achievement (Jimerson & Ferguson, 2007). The learning gain they get from repeating a grade is already gone by the end of the following grade (Vandecandelaere, Vanlaar, Goos, De Fraine, & Van Damme, 2013). Yet, despite its ubiquity and questionable utility, overage students are rarely the focus of research on education in low-income countries (Taniguchi, 2015).

1.4.2 Language of Instruction and Home Language of Students

For many students, the language of instruction is different from the language spoken at home. Furthermore, this is sometimes also different from the language spoken across the community. In these complicated cases, education requires not only the acquisition of the curriculum's content but also the acquisition of the language through which the curriculum is conveyed. Because of this, there is broad agreement in the field that when the home language is different from the language of instruction in school, students' literacy attainments could be slowed. A 26-year review of the literature on children's literacy attainments in low- to middle-income countries conducted by Nag et al. (2018) helped provide the data to support this assertion. "Africa is one of the most linguistically diverse continents, accounting for 30% of the world's languages while having only 15% of its population" (UNESCO, 2019).

In contrast, Europe, which has just over 10% of the world's population, only accounts for 4% of its languages (Clegg & Simpson, 2016). While there is widespread consensus that

education should be delivered to students in the language spoken at home, this standard is not yet in place for hundreds of millions of students (UNESCO, 1953; Benson, 2002; Brock-Utne, 2005; Clegg, 2005; UNESCO GMR, 2015). This lack of instruction in a known language limits students' ability to develop foundations for learning. By one estimate, as much as 40% of the global population does not have access to an education in a language they speak or understand (Walter and Benson, 2012). These challenges are most prevalent in regions where linguistic diversity is greatest, such as sub-Saharan Africa, Asia, and the Pacific (UNDP, 2004). This is further supported by UNESCO's research which notes that, in many countries, students are taught and take tests in languages they do not speak at home, hindering the early acquisition of vitally essential literacy skills.

As a multilingual country, Sierra Leone is particularly vulnerable to these hurdles around the language of instruction. Sierra Leone is a country with both indigenous and exogenous languages, the latter group including English, French, and Arabic. While Sierra Leone's instruction medium is English, there is renewed interest in mother tongue education in Sierra Leone, as in many other former African colonies. Mother tongue is 'the language which a person acquires in early years and which normally becomes their natural instrument of thought and communication' (UNESCO, 1953, p.46). This new interest is often grounded in UNESCO's model of mother tongue literacy (UNESCO, 1953) that has been critically re-examined by Tabouret-Keller et al. (1997). In Sierra Leone, the 1991 National Constitution and the New Education Policy emphasized basic education and the teaching of Sierra Leonean languages, both as media of instruction during the first three years of primary schooling and as subjects of study from Junior Secondary School (JSS) upwards. Despite this, since independence, the official language policy in Sierra Leone has been to retain 'the use of English in all official

domains as far as possible' (Ministry of Education, 1970). This means English remains the primary language of instruction despite this expressed interest in supporting mother tongue instruction. Research demonstrates that speaking a home language different from the language of instruction negatively correlates with student achievement (Janssen & Crauwels, 2011).

In summary, studies show that a student's proficiency in the language of instruction and particularly their reading abilities are related to overall achievement. This factor becomes more critical when students speak a home language different from the instruction language. This situation is pervasive in a multilanguage context like Sierra Leone.

1.4.3 Home reading environment

Often children's experiences with reading begin at home. Most students become familiar with the nature and functions of written language in advance of their first day in school through watching and participating in reading activities in their communities and homes. While recent research supports that the home literacy environment impacts language and literacy acquisition (Fernald & Weisleder, 2013), the extent to which caregivers and the home literacy environment influence early language and literacy acquisition has been debated throughout the previous century (Clegg, 2005). Theories of acquisition of reading skills continue to advance as the overall trends in science of cognition, behavior, and development are better understood. Research has established the significance of the home reading environment on early reading skills – particularly concerning letter knowledge, phonemic awareness, and vocabulary (Britto & Brooks-Gunn, 2001). Examples of key components within the home reading environment include caregiver involvement in reading activities, particularly caregiver reading practices in the home. These factors have been shown to impact preschool students' development of early reading skills (Dodici et al., 2003). Specifically, the home environment can be an essential

foundation for further growth and development of reading. DeBaryshe et al. (2000) concluded that the home environment allows children to experience positive reading interactions and that these positive interactions can positively impact future reading skills development. These skills include becoming familiar with reading materials, observing the reading activities of others, engaging in collaborative reading and writing behaviors with others, and benefiting from observing the strategies that caregivers use when engaging in reading tasks.

Homes rich in reading materials have been shown to positively impact outcomes for both early reading skills and later reading achievement. The driver of home conditions are often parents or other caregivers. Because of this, caregivers are vital components in a child's ability to develop reading skills. In addition, home environments that demonstrate a rich, complex language setting, particularly through regular, ongoing conversation, are significantly associated with an increase in a child's language and reading development. Rush (1999) researched the elements of these positive home-based reading activities in low-income environments. Two preschool classes consisting of thirty-nine students, as well as their caregivers, took part in the study. In this study, observations of the home environment were conducted using the CIRCLE-2 assessment. In addition, the Simple Formative Reading Survey and the Peabody Picture Vocabulary Test-revised assessment were also used to address the research questions. This study's results revealed that high reading-related activity rates were positively associated with better reading outcomes (Rush, 1999).

Involvement in other home-based reading practices has also exhibited positive outcomes for early reading skills development. These activities include reciting nursery rhymes, telling stories, having conversations at dinner, and playing reading-based games (Skibbe et al., 2008;

Weigel et al., 2006). In addition, book reading is a reading activity that robustly predicts future reading skills (Bus et al., 1995).

Furthermore, Burgess et al. (2002) established that regular, recurring joint book reading between caregivers and children influenced oral language, letter-sound knowledge, phonological awareness, and decoding skills. This study focused on communal reading behaviors between caregivers and children. It determined that joint reading behavior significantly affected later academic abilities of preschool children (Bracken & Fischel, 2008). This study's research questions centered on the connections between home-based reading behaviors and early reading and language skills development. Specifically, the study focused on how a family's reading behavior influences a child's reading ability. This study consisted of two hundred thirty-three students attending a full-day preschool program in the New York area and examined the reading interactions between the caregiver and child through the frequency of shared reading and the age when shared book reading began. The study found these activities were significantly correlated with multiple emergent reading skills, particularly related to story and print concepts (Bracken & Fischel, 2008).

Not only has the volume of exposure to reading been shown to be important for future reading skills development, but the frequency of high-quality caregiver instructional interactions also proved a powerful influence (Sonnenschein & Munsterman, 2002). The quality of book reading can also be understood in relation to three core activities: the caregiver encourages the child to discuss the illustrations during reading, promotes the extension of thought and provides corrective feedback to the child, and adapts to the child's developmental level. Ultimately, these actions allow the child to become the storyteller of the book. This is in opposition to a more standard approach where the child is thought to be passively learning through listening to a

caregiver read the book (Mol et al., 2008). Ultimately this study found that the reading activities between caregiver and child explain around 8% of the variance in early reading skills (Hood et al., 2008).

Blom-Hoffman et al. (206) evaluated the effects of caregiver use of this reading technique on children's vocabulary. Utilizing a randomized experimental design, the researchers constructed a sample of 18 children with a mean age of 4. Their caregivers also participated in the study. The caregivers in the experimental group were taught supportive reading techniques through video instruction. They were also given handouts outlining the methods as a takeaway reminder for everyday use. Data collection was done via video observations which were scored with regard to reading techniques demonstrated in the interactions between the caregiver and child. The results show that literacy skills increased with the caregivers' use of supportive reading strategies (Blom-Hoffman et al., 2006).

Cline and Edwards (2013) study of eighty-one children under the age of three, along with their caregivers, showed that caregivers' instructional quality positively impacts young children's emergent reading outcomes. The investigators assessed recorded reading exchanges between the caregiver and the child. The exchange was evaluated on questions asked to the child, the type of constructive feedback by caregivers, conversations related to the book, and direct reading. The results determined that these particular strategies positively impacted a child's reading skills development (Cline & Edwards, 2013).

Foster et al. (2005) developed a study to examine the effects of preschool children's home environment on early reading skills acquisition. The researchers utilized structural equation modeling to explore the mediating impact of the home environment on socioeconomic status and students' emergent reading competence. This study consisted of 421 preschool

students, as well as caregiver interviews. For the purposes of the study, the home environment encompassed general reading activities with the child and enriching literacy experiences completed with the child. The study found that the home environment mediates the relationship between socioeconomic status and emergent reading skills (Foster et al., 2005).

According to Sonnenschein and Munsterman (2002), children who have opportunities to interact with print materials before entering school show an increased interest in reading. Another study by Bennett et al. (2002) showed that children's reading motivation was positively influenced by access to libraries and their caregiver's participation in read-aloud activities. The study also found that a child's reading motivation was increased by recalling nursery rhyme and telling stories. These associations were still significant over 1 year later. While the links between home factors and reading skills development have been explored globally, limited research has been conducted focusing on Sierra Leone.

1.5 Education and Reading Instruction in Sierra Leone

While an LMIC, Sierra Leone has demonstrated the potential for economic growth and positive health and education developments. From 1991 to 2002, Sierra Leone endured a violent and destructive civil war, leading to over 50,000 deaths and two million people displaced. Additionally, civil infrastructure was left in ruins, including the education system. As the country has stabilized in recent years, the government has firmly focused on improving the quality of education – particularly in primary grade learning. Due to this, enrolment rates in Sierra Leone have improved over the last few years, but challenges of low and inequitable access to school persist across the system. At the same time, most children enter primary school; however, many drop out of school before completing a basic education cycle.

Currently, most pupils in Sierra Leone are not learning how to read in school (UNICEF, 2014). While the Government of Sierra Leone (GoSL) declared basic education "free and compulsory" with the Education Act of 2004 (*The Education Act*, 2004), according to the most recent Demographic and Health Survey, only 31.9% of males over six and 23.7% of females have completed primary school or higher (Ministry of Health and Sanitation, 2013). Furthermore, Sierra Leone's education system was devastated by the 2014–2015 Ebola virus outbreak; schools closed for more than 9 months, resulting in nearly 1 year of lost schooling (GoSL, 2018). Schooling was again closed for 8 months due to the COVID-19 pandemic.

Despite these challenges, the GoSL has committed to increasing its investment in the education sector. For example, it allocated 21% of the national budget to support the GoSL's Free Education Program launch in August 2018. The program provides free education from pre-primary through secondary school and strengthens schools' infrastructure, supply chains, and services (State House Media and Communications Unit, 2018). In addition, the Ministry of Basic and Senior Secondary Education's 2018–2020 Education Strategy aimed to increase access, equity, and completion rates; improve pupils' education quality and relevance; and strengthen the education system. Critical interventions of the robust strategy include bolstering the national school feeding program, upgrading school infrastructure through maintenance or construction, improving teaching and learning materials in the classroom, and investing in teachers' skills and motivation (GoSL, 2018).

In support of the GoSL's Education Sector Plan, Catholic Relief Services implemented the All Pikin for Learn (APFL) project in northern Sierra Leone since 2008. APFL, funded by the U.S. Department of Agriculture's McGovern-Dole Food for Education program, strives to reduce hunger and improve literacy and primary education. McGovern-Dole projects provide

school meals, teacher training, and other support activities to boost school enrollment and academic performance.

1.6 Aims and Research Questions

This dissertation aims to review the relationship between student literacy achievement in Sierra Leone and social-contextual factors, focusing on the home environment. As demonstrated in the literature review, there are many strengths and weaknesses of EGRA as an assessment. Particularly in relationship to the quality of the data produced. Due to this, before examining the relationship between critical social-contextual factors and EGRA in this cohort, a review of the quality of the instrument will be undertaken. Correlations between student achievement on the EGRA and the key social-contextual factors of age, language, and home reading environment will be explored.

The following research questions are evaluated:

RQ1: What do descriptive statistics tell us about students' reading levels within the 2019 Grade 2 Sierra Leone EGRA cohort? What impact did zero scores have on the analysis?

RQ2: What conclusions about the quality of the assessment can be drawn from an examination of the psychometric properties of the 2019 Grade 2 Sierra Leone EGRA through a Rasch Measurement Framework?

RQ3: How to social-contextual factors such as age, language, and home reading environment linked to early grades reading achievement within the 2019 Grade 2 Sierra Leone EGRA cohort?

2 Methods

As discussed in the introduction, the EGRA has been developed and implemented in many countries, including Sierra Leone, to understand students' emerging reading skills better.

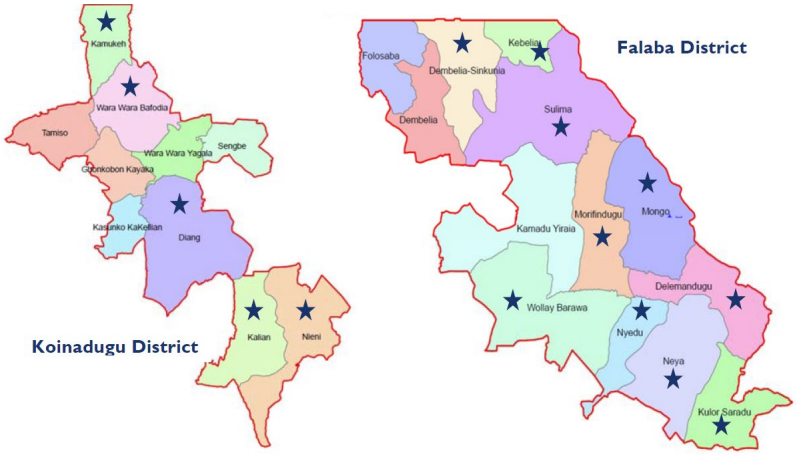
The literature review discusses that literacy skills are impacted by social-contextual factors, particularly the home environment. By reviewing the quality of the assessment used, as well as the social-contextual factors influencing student performance, a more robust understanding of the target population can be gained. The study participants are described below, as well as a description of the tools used and the ethical issues considered. The assessment structure is described in the materials section, followed by information on the type of data available to use in the analysis and details of the data preparation procedure. Finally, a brief outline of the Rasch model used is given, with details of model evaluation.

2.1 Participants

The data for the present study consisted of 682 Grade 2 students from 15 schools across 15 chiefdoms in northern Sierra Leone in June 2019 (See Figure 1). Each subtask in the analysis may present data for a slightly different number of students due to data cleaning issues. For each student, EGRA data and relevant background characteristics – such as gender and geography - were collected using a suite of quantitative and qualitative tools. These tools included the EGRA, and surveys conducted with pupils, teachers, and community members.

Figure 1:

Intervention Chiefdoms within the Koinadugu and Falaba Districts



2.2 Ethical Issues

The UNEG Ethical Guidelines for Evaluations and the 'UNICEF Procedure for Ethical Standards in Research, Evaluation, Data Collection, and Analysis' were followed during the initial data collection. Due to the nature of the secondary data used in this study, no CUREC was submitted. All permissions were sought and secured from the originators of the data collection. There is no expected physical, psychological, social, or legal risks to respondents. The main risk is a breach of confidentiality which was mitigated by ensuring that all analysis was done using anonymized data stored securely using encrypted, password-protected files.

2.3 Materials

The tools include an EGRA and pupil survey and school-based stakeholder surveys. The assessment contained seven untimed subtasks administered in English: alphabet naming, phonemic awareness, familiar word reading, invented word reading, reading passage, reading comprehension, and listening comprehension. Table 2 provides a summary of the subtasks.

Table 2

EGRA Subtasks Utilized in this Assessment

Subtask	Core reading skill	Subtask description
Alphabet naming	Alphabet knowledge	Provide the name of 51 letters presented in both uppercase and lowercase in random order.
Phonemic awareness	Phonemic awareness	Identify the words represented by 10 pictures and give the sound of the first letter of each word represented.
Familiar word reading	Word recognition	Read 40 familiar words randomly ordered and drawn from a list of frequent words.
Invented word reading	Decoding	Make letter-sound correspondences through the reading of 25 simple nonsense words.
Reading passage	Decoding and reading	Read a short, grade-appropriate passage of 36 words with accuracy and little effort.
Reading comprehension	Reading comprehension	Respond correctly to five questions about the passage read in the previous subtask, including four literal questions and one inferential question.
Listening comprehension	Listening comprehension and oral language	Listen to a text the enumerator reads aloud and respond correctly to four questions about the text, including three literal questions and one inferential question.

For a comprehensive picture of the students' home environment, numerous data points were collected at the school level (see Table 3).

Table 3

School-Based Surveys and Observation Checklists

Tool	Types of information collected
Student Survey	Socio-economic demographics at home, household literacy levels, learner school history.
Teacher survey	Levels of teacher certification; in-service training and coaching; knowledge and use of teaching techniques; motivating factors; reading materials present at home, satisfaction with the APFL project.
Head teacher survey	MBSSE status; enrollment and attendance data; teacher training, attendance, retention information; school infrastructure details; teaching and learning materials available; school activities and support structures.

The student surveys included closed and open-ended questions with in-built marking categories designed to standardize responses. The tool covered several domains, including socio-economic demographics at home, household literacy levels, and learner school history. The head teacher and teachers in each school who were observed were interviewed using tools that comprised open-ended questions across a variety of domains related to teaching and learning. Specific questions focused on the inputs provided to date regarding teacher training, reading materials available at school and home, and content knowledge.

2.4 Procedure

A two-stage cluster sampling approach was used. Attributes such as district location—were accounted for within the sample. First, schools were randomly selected as clusters. For the second sampling stage, 10 pupils were randomly selected from those present in Class 2 at each sample school. A list of appropriate replacement schools was also created if the original sample schools were unavailable or difficult for enumerators to reach. The target sample size was 70 schools and 700 pupils—10 pupils, five girls, and five boys, from each school.

Enumerators entered learning assessments, student, and teacher/head teacher interview responses directly on tablets. The data was uploaded every evening to the server. Throughout data collection, the field coordinator reviewed the uploaded data to ensure consistency with the detailed field report and ensure that no assessments were missing. In case of any irregularities, teams were expected to provide a valid explanation or to return to schools to complete any missing assessments. At the end of the data collection process, all the data had been uploaded and reviewed again before data analysis took place. In addition, the field coordinator visited multiple schools to conduct on-site spot checks and troubleshoot any issues encountered by teams in the field. Communication with the enumerator teams was maintained through a WhatsApp group comprised of team supervisors; this allowed for broader communication and faster responsiveness when issues arose in the field. However, many enumerator teams could not upload their data electronically daily due to severe connectivity issues in the communities. Sometimes, data could not be uploaded until the teams returned to Freetown. This impeded the field and program coordinator's real-time data-tracking activities.

The analysis used sampling weights to produce more representative estimates in the sample of pupils. Although random sampling does not acknowledge that some pupils have a lower probability of being selected when they represent smaller subgroups within the population, sampling weights allow the analysts to account for these differences in probabilities. Weights were computed using background data available from each school in the sample populations, including the number of Class 2 classrooms at the school and the number of pupils in each classroom. This information was collected via the head teacher survey. Weights were applied when analyzing the reading assessment and survey results.

2.5 Data Analysis

Descriptive statistics will be constructed for each subtask to begin exploring the data. Descriptive statistics are an analysis of data that helps describe the data in a meaningful way such that patterns might emerge from the data. For this data, means, standard deviations, and correlations will be explored. In addition to descriptive statistics, Rasch modeling was used. Rasch methods are not commonly applied to EGRA data. Still, they have been chosen here because they offer an alternative method to understand the quality of the assessment and student achievement. The models describe a probabilistic mapping of the observed pattern of student-item responses from an assessment to a latent trait (θ) that is assumed to represent the test's ability to assess. Both student abilities and item difficulties can then be described in terms of their position on the latent trait, rather than in terms of mean or total marks.

The Rasch model utilizes the principles of IRT to analyze assessment data. Rasch determines the probability of a person responding correctly or incorrectly to each item on an assessment on which a person answers questions either right or wrong. The core principles of the Rasch model are: 1) that a higher-ability person will have a greater probability of getting an item correct than a lower-ability person and 2) that a given person will have a higher likelihood of getting an easier item correct than getting a more challenging item correct. The relationship between these two core principles serves as the model's fundamental basis. Ultimately, the Rasch model seeks to measure a latent unidimensional trait. A latent unidimensional trait can be understood as an underlying construct – meaning a construct that cannot be measured directly - that can be thought of as lying on a single dimension – meaning in this case that it can be thought of in terms of more or less (Luppescu & Ehrlich, 2012).

The Rasch model is the simplest latent trait model since it operates as a one-parameter IRT model. In Rasch, the latent trait is considered as a single dimension along which items can be located in terms of their difficulty and people can also be located in terms of their ability. Additionally, the Rasch model is a probabilistic model that estimates the probability of answering the item correctly as a logistic function of the difference between the person's ability and the item's difficulty. To do this, the Rasch model establishes an interval scale of scores for both the item's difficulty and the person's ability which are scaled in logits. The Rasch model has the property of specific objectivity, which may be interpreted as sample independence. This is its advantage over other IRT models. Specifically, the Rasch model (1960) for dichotomously scored data describes a student's ability or an item's difficulty as a function of the responses on the exam paper. The items' difficulties are then represented by a set of parameters (βi) that give their location on the latent trait (θ), higher values of βi indicate more difficult items. Student abilities are also described as a location on the latent trait. The probability of a student giving a correct response increases as their level on the trait increases (i.e., higher values of θ indicate higher student abilities).

$$P(X_{ni} = 1) = \frac{e^{\theta_n - \beta_i}}{1 + e^{\theta_n - \beta_i}}$$

Where:

Xni is the response (0 or 1) of the nth student on the ith item

θn is a parameter describing the ability of the nth student, $-\infty < \theta n < \infty$

βi is a parameter describing the difficulty of the ith item, $-\infty < \beta i < \infty$.

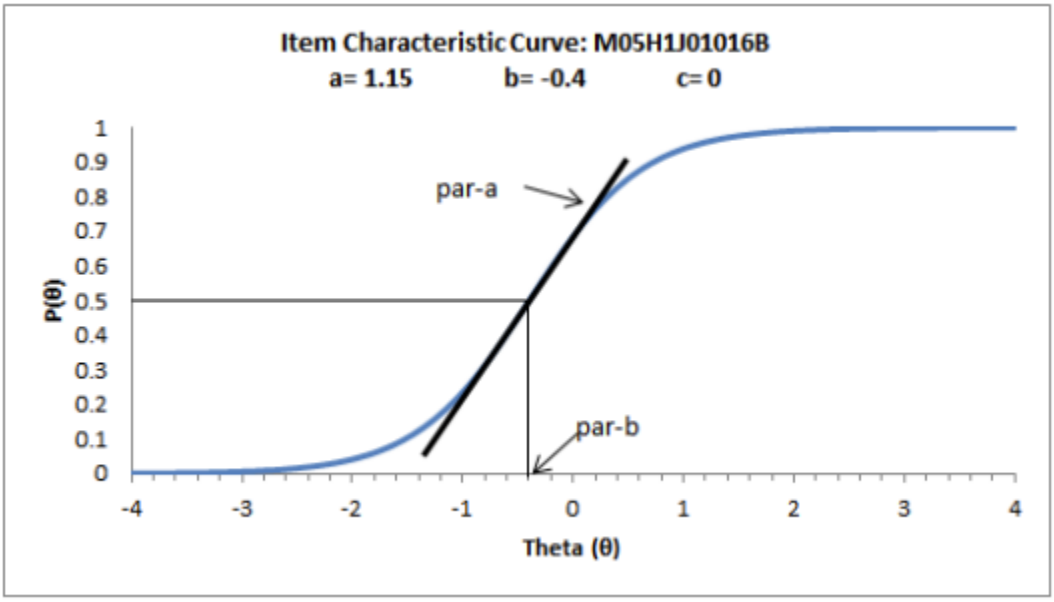
It can be shown that the log-odds (logit) of a correct response by student *n* on item *i* is given by the difference in their ability and the item's difficulty: $\theta n - \beta i$. When a student's ability

is equal to an item's difficulty, the probability of a correct answer is .5. When their ability is greater than the item difficulty, the probability of a correct answer will be higher, and vice versa. For each item, the relationship between a student's ability and the chances of answering the question correctly can be illustrated graphically. These curves are known as item characteristic curves (ICC), where the latent trait of ability, represented by θ on the x-axis, has a value between -3.0 and 3.0, and the probability of answering an item correctly, P(θ) on the y-axis, is between 0.0 to 1.0. An example can be seen in Figure 2 below.

An examination of the ICC can help identify a misfit. ICCs visualize the probability of success on the item for each possible location of a person on the continuum. The location of each item is the point at which a person would have a 50% chance of being successful on the item. If the data fit the model, then the mean of the set of person estimates in each class interval should be close to the theoretical curve; in other words, the dots should follow the line. Items with a good fit tend to show each of the group plots lying on the curve. Those with plots that were steeper than the curve would be considered to be over-discriminating, and those flatter than the curve under discriminating (Andrich & Marais, 2019)

Figure 2:

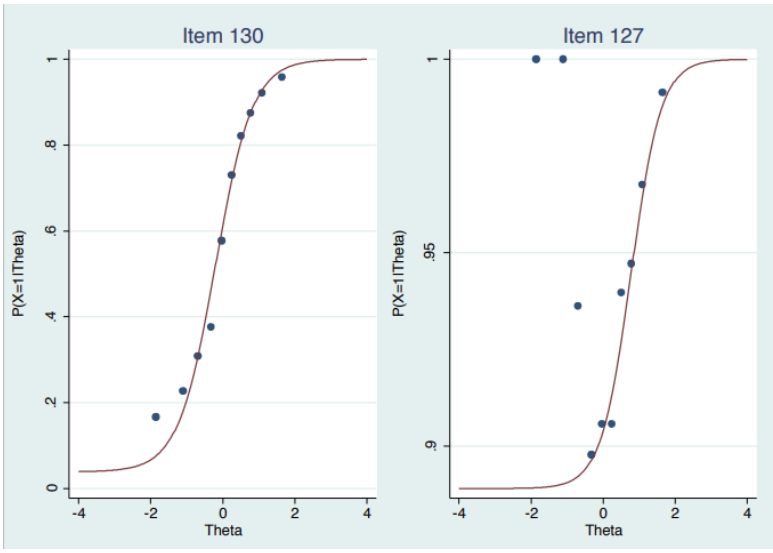
Example ICC Curve



The item difficulty can be seen in (b-parameter) in the example above. The b-parameter shows the item location on an ability scale, which represents a probability of 50% that a pupil will get the item correct. This means that higher values within the b-parameter represent more difficult items. The a-parameter shows the slope of the ICC. The slope represents how well an item differentiates between high-performing and low-performing test takers. The steeper the slope, the better the item is at separating test takers of different ability levels. Below, Figure 3 offers an example ICC curve demonstrating items with good and bad fit.

Figure 3:

Example ICC curves Demonstrating Good and Bad Fit



Rasch models also enable the estimation of standard errors for the item parameters and the student ability estimates (Hambleton et al., 1991), which facilitates the evaluation of the fit of item and student estimates. Each of the models described above assumes that the latent trait is unidimensional; that is, there is no other significant dimension captured by the test other than the ability the test was designed to measure. It is also assumed that responses to the items are independent; after considering student ability, there is no relationship between the response to pairs of items (known as local independence).

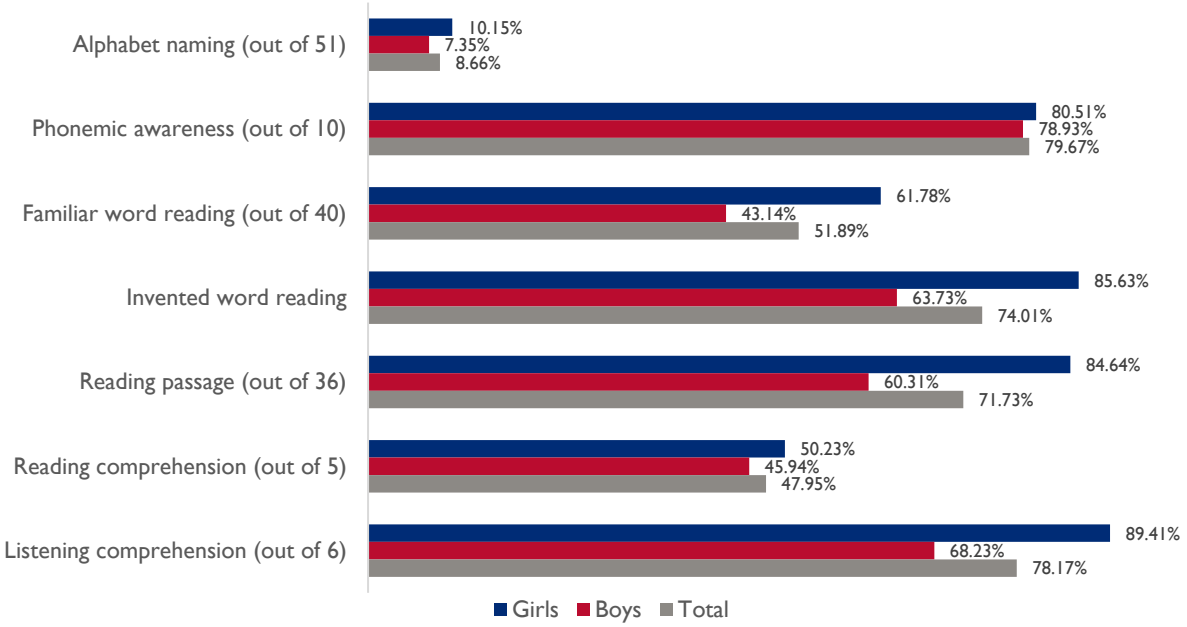
3 Results

3.1 Descriptive Statistics

The proportions of pupils who did not answer a single item correctly across subtasks—known as zero scores—are presented in Figure 5. The proportion of pupils receiving zero scores was lowest on the alphabet naming subtask (8.66%) and highest on the phonemic awareness subtask (79.67%). Across all subtasks, boys had a lower proportion of zero scores than did girls.

Figure 4

Percentage of Pupils Receiving Zero Scores by Sex



Note: Girls (n=334); Boys (n=348); Total (N=682)

Mean scores for other reading assessment subtasks are presented in the following section to understand pupils' reading performance better. Statistical significance tests were performed to analyze the difference in mean scores between boys and girls; statistically significant differences are noted under each table.

3.1.1. Alphabet Naming

In the alphabet naming subtask, enumerators presented pupils with a grid of 51 letters in uppercase and lowercase and asked students to say the name of the letter.[1] The alphabet naming subtask measures pupils' knowledge of letters of the alphabet and their ability to recognize the graphemic features of each letter.

Baseline results for the alphabet naming subtask are presented in Table 5. On average, pupils named 36.57 letters correctly out of 51. Statistically, boys had significantly higher mean scores than girls; boys, on average, named nearly five more letters than girls.

Table 5

Alphabet Naming Mean Scores by Sex

Sex	*N*	Mean score	Standard error
Girls	334	34.06	.94
Boys	348	38.79**	.83
Total	682	36.57	.63

**Boys' scores are significantly higher than girls' scores at the $p < .01$ level.

3.1.2 Phonemic Awareness

For the phonemic awareness subtask, enumerators provided pupils with clip art pictures of 10 common objects and read the name of the object out loud to the pupils. Pupils were asked to say the initial sound of the object's name. The phonemic awareness subtask measures pupils' awareness of phonemes and their ability to distinguish among multiple phonemes.

Baseline results for the phonemic awareness subtask are presented in Table 7. Out of 10 possible items, pupils correctly identify the initial sound of 1.06 items on average. Again, there was no significant difference between girls' and boys' scores.

Table 1

Phonemic Awareness Mean Scores by Sex

Sex	*N*	Mean score	Standard error
Girls	334	0.94	.13
Boys	348	1.16	.14
Total	682	1.06	.09

Note. Scores are correct out of 10.

3.1.3 Familiar Word Reading

The familiar word reading subtask consisted of a list of 40 familiar words that were presented to pupils in a grid. Pupils were asked to read as many of the familiar words as they could out loud.[2] Familiar word reading measures pupils' sight-word recognition and decoding skills.

Baseline results for the familiar word reading subtask are presented in Table 8. Pupils correctly read an average of 11.60 familiar words at baseline. Boys read nearly twice as many words correctly as girls—14.98 familiar words versus 7.78 familiar words, respectively. The difference between girls' and boys' mean scores was statistically significant.

Table 2

Familiar Word Reading Mean Scores by Sex

Sex	*N*	Mean score	Standard error
Girls	334	7.78	.67
Boys	348	14.98 [a]	.80
Total	682	11.60	.54

Note: Scores are correct out of 40.

[a] Boys' scores are significantly higher than girls' scores at the $p < .01$ level.

[2]. The familiar words on the assessment were rerandomized within lines. One item—"play"—appeared twice in the grid.

3.1.4 Invented Word Reading

For the invented word reading subtask, pupils were presented with a grid of 25 made-up words that follow the phonological and spelling rules of English but are not actual words in the language. Next, enumerators asked pupils to read aloud as many non-words as they could.[3] Invented word reading measures pupils' decoding skills.

Baseline results for the invented word reading subtask are presented in Table 9. Out of 25 items, pupils correctly read 3.31 invented words on average. Boys outperformed girls, reading 4.88 invented words correctly compared with 1.54, on average; this difference was statistically significant.

Table 3

Invented Word Reading Mean Scores by Sex

Sex	*N*	Mean score	Standard error
Girls	334	1.54	.24
Boys	348	4.88 [a]	.38
Total	682	3.31	.24

Note: Scores are correct out of 25.

[a] Boys' scores are significantly higher than girls' scores at the $p < .01$ level.

3.1.5 Reading Passage and Reading Comprehension

For the reading passage and reading comprehension subtasks, pupils were presented with a short story of 41 words and asked to read as much of the story aloud as they could. After finishing, enumerators read five comprehension questions—four direct and one inferential—out

loud to pupils to test their understanding of the story's content.[4] These two subtasks measure decoding and reading comprehension.

Baseline results for the reading passage subtask are presented in Table 10. From a short story of 41 words, pupils correctly read 7.59 words on average. While boys correctly read about 26% of the words (10.89 words), girls only read about 9% (3.87 words) correctly. The difference in mean scores between girls and boys was statistically significant.

Table 4

Reading Passage Mean Scores by Sex

Sex	*N*	Mean score	Standard error
Girls	334	3.87	.54
Boys	348	10.89 [a]	.76
Total	682	7.59	.49

Note: Scores are correct out of 41.

[a] Boys' scores are significantly higher than girls' scores at the $p < .01$ level.

Baseline mean scores for the reading comprehension subtask are presented in Table 11. Overall, pupils were able to answer less than one reading comprehension question correctly at baseline. In addition, boys had significantly higher mean scores than girls on the reading comprehension subtask.

Table 5

Reading Comprehension Mean Scores by Sex

Sex	*N*	Mean score	Standard error
Girls	334	0.31	.05
Boys	348	0.86 [a]	.08
Total	682	0.60	.05

Note: Scores are correct out of 5.

[a] Boys' scores are significantly higher than girls' scores at the $p < 0.01$ level.

The distribution of pupils able to answer reading comprehension questions correctly is detailed in Table 12. Overall, nearly four out of five pupils (78.17%) were unable to answer a single reading comprehension correctly.

Table 6

Distribution of Correct Reading Comprehension Questions by Sex

Number of questions correct	Girls	Boys	Total
0	89.41%	68.23%	78.17%
1	1.19%	7.22%	4.39%
2	2.69%	7.05%	5.00%
3	3.60%	9.73%	6.85%
4	1.95%	3.56%	2.80%
5	1.17%	4.21%	2.78%

3.1.6 Listening Comprehension

The listening comprehension subtask consists of a short story of 40 words, which was read out loud by the enumerator to the pupils. The enumerator then asked the pupil four comprehension questions related to the story—three direct and one inferential. Listening comprehension measures pupils' overall oral language comprehension and vocabulary. The listening comprehension subtask complements the reading passage and comprehension subtasks,

as it enables a better understanding of whether pupils' comprehension difficulties result from reading skills or overall language comprehension.

Baseline results for the listening comprehension subtask are presented in Table 13. Out of a possible four questions, pupils correctly answered, on average, 1.06 questions. Again, there was no statistically significant difference between boys' and girls' performance.

Table 7

Listening Comprehension Mean Scores by Sex

Sex	*N*	Mean score	Standard error
Girls	334	1.00	.07
Boys	348	1.11	.07
Total	682	1.06	.05

Note. Scores are correct out of 4.

3.2 Rasch Model

The sections below summarize the results from the diagnostics tests carried out on the pupil test data to assess its fit with the Rasch model. The fit-of-model must be evaluated to draw valid inferences from an item response theory (IRT) model, such as Rasch. Model misfit indicates that at least one, or several, of the model assumptions have been violated (Köhler et al., 2017). Therefore, the following fit statistics have been inspected to assess the fit of the data to the Rasch model. These figures can be seen in Table 14 below.

Table 14

Summary of Fit Statistics

Analysis	Item fit residuals		Person fit residuals		Item–trait interaction		PSI
	Mean	*SD*	Mean	*SD*	Chi-square (DoF)	*P*	
	−1.08	4.50 [a]	−0.19	0.88*	352.46 (99) [a]	0.00	0.833

SD = standard deviation. DoF = degrees of freedom. PSI = Person Separation Index.

Three fit statistics can be examined to establish model fit. Two were fit residual statistics—person fit and item fit. These represent the residuals between the expected estimate and actual values for each person-item, summed overall items for each person (person fit), and overall persons for each item (item fit). The Person Separation Index (PSI) is the third statistic. PSI evaluates the scale's internal consistency reliability and the measure's ability to discriminate among persons with different levels of the underlying trait (Lamoureux et al., 2006). Finally, the item fit residual statistic provides evidence of the fit of the data to the model from the perspective of the items. In this case, the item fit residuals for the data demonstrated a mean across all items of −1.08 with a standard deviation of 4.50 (Table 14). This high-positive-standard-deviation fit residual indicates that the test exhibits strong signs of under-discrimination.

Similar to item fit residuals, person fit residuals are constructed for each person and then further transformed across items to be similar to a normal standard deviation. As with item fit residuals, the mean and standard deviation will ideally be close to 0 and 1, respectively (Andrich, 1988). In this case, the person fit residual mean is −0.19, and the person fit residual standard deviation is 0.88

Item–trait interaction is used to assess the invariance across the trait. This is reported as a chi-square (Wright & Stone, 1979). When the chi-square value is significant, this indicates that there may be variance across the trait for the hierarchical ordering of the items. This would suggest that the required property of invariance had been compromised. In this case, the chi-square was not significant, indicating a possible model misfit (Table 14).

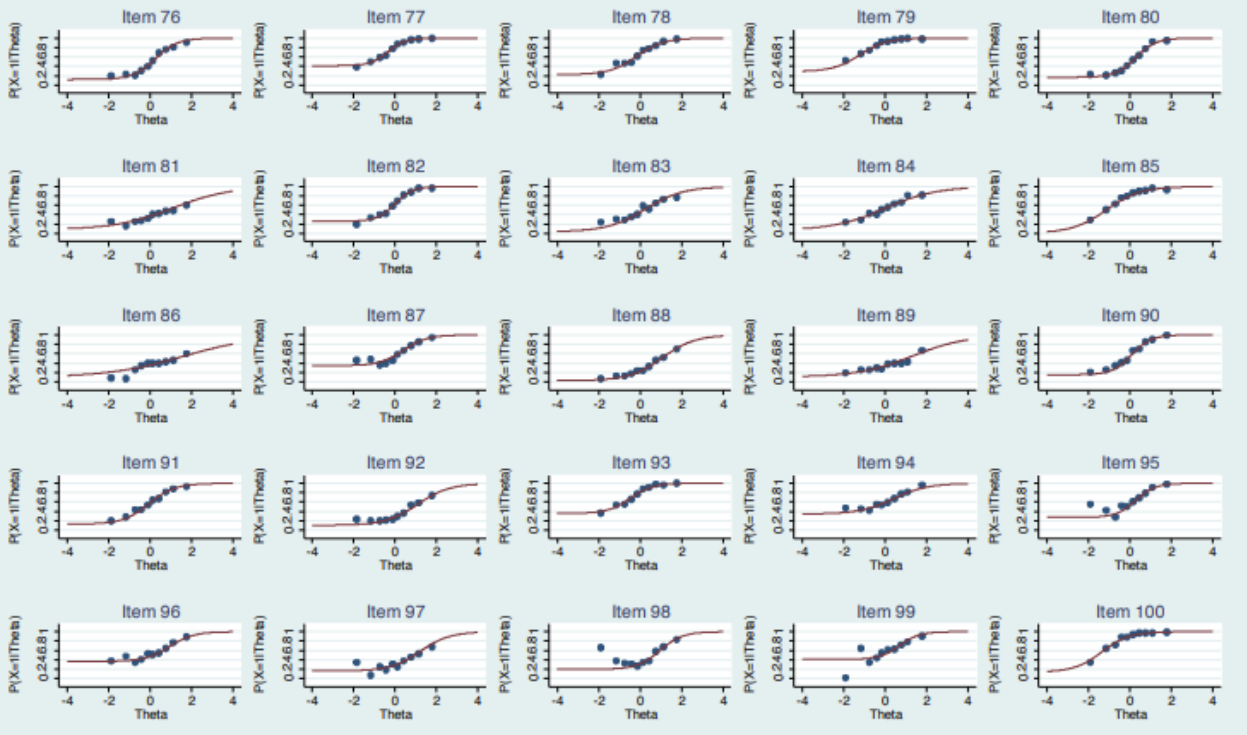
The PSI shows the ability of the construct to discriminate among the students. A value of 0.7 is conventionally considered the minimum acceptable level of the PSI. The PSI determines

the number of groups of students who can be statistically differentiated from each other (Wright, 1996). The PSI is also an indicator of the reliability of the fit characteristics (Wright, 1996). A lower PSI indicates less reliability. This data set shows a PSI of 0.79, indicating group analysis reliability (Table 14).

In addition, the ICC's for the essential items were visually examined for fit. As seen in Figure 5 below, the items generally showed a good fit across the assessment.

Figure 5:

ICC for Key Assessment Items



While most items show a reasonable fit, item 86, item 95, item 96, and item 99 indicate some misfit. These items are Reading comprehension items with a very small number of respondents.

3.3 Social-Contextual Factors

3.3.1 Age, Grade Level, and Links to Early Learning

Many students were, on average, two years older than they should be in grade 2. Additionally, only 1 in 4 learners attended pre-primary school before primary school. All students who attended pre-primary school demonstrated at least some skills on the EGRA assessment – meaning they did not score a zero on all tasks. Students who attended pre-primary were notably stronger in their pre-reading skills. A substantial number of grade 2 learners (about 16-18 percent) are aged 9-10 when they should be 7-8 years. Many underage learners are also enrolled, notably children of 6-7 years who should still be in grade 1. The distribution of student age can be seen in the table below:

Table 15:

Student distribution by age group for Grade 2 Learners

Age	Percentage of Students Sampled
6 years old	5.8
7 years old	23.5
8 years old	31.3
9 years old	19.3
10 years old	12.9
11 years old	3.8
12 years old	2.7
13 or more years old	0.8

3.3.2 Home Language and Language of Instruction

For many students, English is their second or third language. English and Krio are largely spoken in equal measure in school during instruction. However, English is the only reported language of instruction in national documents and the only language found in textbooks. About 30 percent of learners reported speaking their local language in school during class instruction. Speaking English at school positively correlates with learning achievement, while speaking Krio or another local language has the opposite effect. Seventy-five percent of learners reported speaking their local language at home with family, while another 50 percent use Krio. Very few students reported speaking English at home, meaning that most learners only get exposure to this language of instruction at school. Speaking English or Krio at home positively correlates to learner performance, demonstrating they are a driver of better results. This has an impact on literacy development.

Table 16:

Student Distribution by Language Spoken at Home and at School

Language	Spoken at School	Spoken at Home
English	65%	1%
Krio	56%	49%
Local language	28%	75%

Table 17:

Significance of Home Language by Subtask

Subtask	Home Language - English			Home Language - Krio			Home Language - Other language		
	Coef	Std. Err.	p-val	Coef.	Std. Err.	p-val	Coef	Std. Err.	p-val
Alphabet Naming	-1.1	2.01	0.59	1.2	0.54	0.031*	0.7	0.62	0.29
Phonemic awareness	9.3	1.68	0.001**	2.2	0.46	0.001**	-2.3	0.50	0.001**
Familiar words	14.9	1.76	0.001**	3.6	0.42	0.001**	-3.8	0.46	0.001**
Non-words	18.3	2.42	0.001**	5.0	0.63	0.001**	-6.4	0.70	0.001**
Reading comp	24.3	4.04	0.001**	5.0	1.03	0.001**	-6.9	1.14	0.001**
Listen comp	-0.1	0.10	0.47	0.0	0.02	0.89	0.0	0.03	0.26

* Notes p-value of less than 0.05 ** Notes p-value of less than 0.01

3.3.3: Home Literacy Environment

About 50 percent of learners' fathers and 40 percent of their mothers are literate. Seventy-five percent of learners reported reading and studying at home with a range of other family members and friends. Most learners read with their siblings. Just over 30 percent take books home from school to read, while over 50 percent reported having other reading books at home already, namely storybooks (40 percent). Reading at home had a position correlation with better performance on the EGRA. Interestingly, analysis also showed that mothers reading at home with their children had the greatest statistically significant impact on learner performance,

followed by reading with fathers; reading with a sibling corresponded to negative effects on learning.

Table 18:

Distribution of Students by Who Reads with Them at Home

Who Learner Reads With	Percentage
Reads with mother	9
Reads with father	13
Reads with sibling	47
Reads with an adult relative at home	12
Reads with friends	4
Reads with no one (reads alone)	10
Reads with another companion	9

Table 19:

Significance of Reading with Family Members by Subtask

Subtask	Reads with mother			Reads with father			Reads with sibling		
	Coef	Std Err	p- val	Coef	Std Err	p-val	Coef.	Std. Err.	p-val
Alphabet Naming	3.8	1.01	0.00	2.7	0.82	0.00	-3.1	0.58	0.00
Phonemic awareness	3.4	1.22	0.011**	0.8	1.03	0.44	-0.4	0.71	0.55
Familiar words	2.2	0.83	0.011**	0.9	0.70	0.19	-1.0	0.49	0.041*
Non-words	2.7	0.86	0.00	2.1	0.69	0.00	-1.7	0.50	0.00
Reading Passage	0.0	0.06	0.45	0.0	0.05	0.56	0.0	0.03	0.50
Reading comp	0.0	0.05	0.99	0.0	0.04	0.48	0.1	0.03	0.021*
Listening comp	0.0	0.07	0.86	0.0	0.06	0.76	-0.1	0.04	0.09

* Notes p-value of less than 0.05 ** Notes p-value of less than 0.01

4 Discussion

This dissertation has evaluated the outcomes of the learning assessment, as well as its quality. After understanding the quality of the assessment, the relationship between social-contextual factors and student reading achievement was examined.

4.1 Summary of Findings

4.1.1 RQ1 - What do descriptive statistics tell us about the reading level of students within the 2019 Grade 2 Sierra Leone EGRA cohort? What impact did zero scores have on the analysis?

Based on the reading assessment, pupils' reading outcomes are low—only 5.58% of Class 2 pupils were able to answer three out of five of the reading comprehension questions. The proportion meeting the threshold was lower for girls (3.11 percent) than for boys (7.77 percent). Pupils performed best on the alphabet naming subtask, but more than half of all pupils struggled with the other subtasks. In general, boys outperformed girls across the subtasks, which indicates a gender-based performance gap. Importantly, low performance on the listening comprehension sub-task suggests that students cannot comprehend a story told to them at a level deemed appropriate, indicating very low English abilities that are foundational for reading acquisition.

More specifically, pupils performed best on the alphabet-naming subtask, but more than half of all pupils struggled with the other subtasks. Notably, 78.17 percent of students could not answer a single listening comprehension question. As noted, this extremely low performance on the listening comprehension subtask indicates that students cannot comprehend a story told to them at an appropriate level, indicating very low English abilities that are foundational for reading acquisition. The low reading comprehension level is further supported by the fact that English was spoken regularly in only 65% of schools and only 1% of homes. The fact that a large segment of students may not have been able to access the content of the assessment due to

language barriers potentially limits the utility and learnings that can be drawn from the data set, as there could be other variables unique to the small number of students who were able to access the assessment.

Additionally, as established by the literature review, phonics are a key component of Sierra Leone's approach to reading acquisition. Despite this focus on phonics, students only correctly identified one out of ten phonemes on average. This indicates a low level of phonemic awareness and may illuminate a potential mismatch between instructional materials and practices with established educational goals. Sierra Leone has established phonics as a key building block toward reading acquisition, which signals that performance on other subtasks will be low.

The familiar word subtask also had a high level of zero scores, demonstrating a statistical difference between boys and girls. While boys correctly read about 26 percent of the words (10.89 words), girls only read about 9 percent (3.87 words) correctly. As established in the literature review, exposure to print can be a crucial component of building familiar word recognition. This may indicate that boys have more regular access to written materials within or outside the classroom.

Overall, the high levels of zero scores mean that there are many students for whom there is no data – making it unclear how many conclusions we can draw about the sampled population. Overall, students performed better on foundational skills – such as letter sound – than on reading comprehension, indicating that they are developing foundational reading skills even if they are not yet reading at grade level. This indicates learners need more practice with beginning literacy

skills that precede whole word reading, including activities that develop their letter-sound knowledge, phonological awareness, and application of the alphabetic principle.

4.1.2 RQ2: What conclusions about the quality of the assessment can be drawn from an examination of the psychometric properties of the 2019 Grade 2 Sierra Leone EGRA through a Rasch Measurement Framework?

Overall the EGRA data represented showed a moderately good fit for Rasch analysis. For the data to satisfy Rasch model requirements: (a) mean is expected to be approximately around zero (can range between 2.5 and −2.5); (b) S.D. should be approximately 1; (c) chi-square value is expected to be small and statistically non-significant, and (d) PSI should be greater than 0.7 to obtain good power for the test of fit. The analysis showed a reasonable fit for the data, but there are possible fit issues. Specifically, the item-fit statistic had a high positive standard deviation of 4.50. This may indicate that the test exhibits strong signs of under-discrimination. In addition, while the person-fit residuals were broadly representative of good fits, the item-trait interaction – as measured by the chi-square, was not significant. This indicates a possible model misfit. Finally, the PSI – 0.79 – indicates group analysis reliability as the conventional minimum standard is often considered 0.7.

Additionally, the individual ICC curves were visually examined for fit. While the items were generally found to be a good fit, some items in the reading comprehension subtask were not found to be a good fit. This may be because there were few respondents to these questions, as the zero scores were relatively high. That said, the EGRA was not designed from the point of view of a future analysis with a Rasch model due to the frequency of zero scores leading to sparse data sets. For example, the proportion of pupils receiving zero scores was lowest on the

alphabet naming subtask (8.66 percent) and highest on the phonemic awareness subtask (79.67 percent).

Furthermore, boys had a lower proportion of zero scores across all subtasks than girls. The data was sparse for some subtasks because few students could complete the task. If data is sparse, then item estimates are only approximate. In addition, a data set full of perfect success (100% correct) or perfect failure (0% correct) does not contribute to the estimation process. Because of this, programs cannot robustly evaluate unexpected responses with sparse data.

Overall, an examination through Rasch shows that the tool may function correctly but, depending on the purpose, may be much too tricky for the students to demonstrate their existing reading skills. For example, if the assessment aims to understand how students are (or are not) progressing toward the benchmark of Grade 2 literacy, then it may function as intended. However, if the test's purpose is to understand better where students are on the path toward reading acquisition, then the test is likely too difficult.

4.1.3 RQ3: How to social-contextual factors such as age, language, and home reading environment linked to early grades reading achievement within the 2019 Grade 2 Sierra Leone EGRA cohort?

Age, grade level, and participation in pre-primary programs are all related to student achievement. For example, students who attended pre-primary programs showed stronger pre-reading skills. This supports the idea that improving the provision of pre-primary instruction could help improve future reading achievement. Additionally, many students were over or underage for Grade 2. In this case, 29.3% of students were underage – potentially representing a

segment of the study population who cannot reasonably be expected to demonstrate Grade 2 skills.

Conversely, 38.78% of students were overaged, demonstrating that students are not making adequate progression. Overall, with only 31.3% of children in the appropriate grade there their maybe competing contextual factors that can account for demonstrated performance on the EGRA. This leads to overcrowding and an early grade bulge that adds cost and wastage to the system – as students are being accessed at levels that may be developmentally inappropriate.

Learner language also had an impact on reading achievement. Only 1 percent of students reported speaking English at home, meaning that most learners only get exposure to this language of instruction at school. Unfortunately, English is only reported as being spoken in school by 65 percent of students – further limiting exposure. This has an impact on literacy development, as less time spent interacting with a language slows children's acquisition of critical foundational skills. Interestingly, speaking English or Krio at home positively correlated to learner performance, demonstrating they are a driver of better results. This may point to the idea that exposure to any language and the local language drives better results. Additionally, the language spoken at school does impact learners' scores. As previously described, when English is the school language, learners perform better, especially in literacy tasks, with an increase in scores. Conversely, learner performance is negatively affected when the school language is Krio or another local language.

While only 50 percent of fathers and 40 percent of mothers were literate, their behaviors in terms of oral language in the home may play a significant role in reading skills acquisition. The language spoken at home plays an even more substantial role than in school: speaking English at home relates to an even more significant increase in reading outcomes. Reading at

home was also an influential factor. Reading at home had a position correlation with better performance on the EGRA. Interestingly, the analysis also showed that mothers reading at home with their children had the greatest statistically significant impact on learner performance, followed by reading with fathers; reading with a sibling corresponded to adverse effects on learning. This is particularly interesting since most students who reported reading at home identified the person they read with as a sibling.

4.3 Challenges, Limitations, and Technical Considerations

Some several challenges and limitations should be considered when considering the findings of this study. Firstly, the generalizability of the results should be carefully considered. This study does not explicitly measure all personal factors that inhibit or support learning. In addition, this study does not account for other underlying factors, e.g., disabilities that might affect children's ability to learn to read. It also does not explore the different social contextual factors that may underlie the traits correlated to better performance. For example, it may be that parents who speak English at home have a higher income level which allows for better healthcare or will enable them to place a higher emphasis on school attendance for their children. Issues like these are a limitation in a location like Sierra Leone, where stunting and widespread health issues are common.

Secondly, the available list of intervention schools provided for sampling was outdated, resulting in a substantial number of schools being replaced because the schools were either not found or not accessible by road/vehicle. While mitigation measures were implemented, this could limit findings' generalizability to students in remote areas.

Thirdly, one-on-one orally administered assessments have enumeration challenges. For example, when training enumerators in cross-regional national data collections, it is often

difficult to control for regional accents or dialects. This means that a student's ability may be over or underrepresenting depending on the enumerator assigned to them for data collection. While a set of standard sounds were agreed on during training, and enumerators were trained to listen for these and to score accordingly on the EGRA tests, this still offers an opportunity for enumerator bias to impact the quality of the data. Where possible, enumerators were assigned to collect data in areas where they spoke the local language; this helped reduce the negative effects accent and dialect may have had on children's pronunciation and on the enumerator's ability to listen for the correct letter sound.

Fourthly, there is an inherent bias in sampling children present on the day of assessment. Pupils' reading assessment results may be biased towards the types of students who attend regularly and may exclude those pupils who are enrolled but do not attend regularly. However, this random method of sampling on the day of the assessment is preferable to sampling pupils in advance, as it may create opportunities for manipulation to have only high performers participate. For example, even with the delivery of the letter from the Ministry and a follow-up phone call from the enumerator team leader, some schools had limited or no teachers or students present in the school on the day of the data collection. This was mitigated, where possible, with repeated contact with the head teachers and rescheduling where necessary.

Fifthly, while there was no direct evidence in this data set of fraud, other assessments enumerated in this area have been subject to data quality issues. Additionally, the data used in this study results from a broader data collection with a long history of broad-scale interventions in the region. As a result, the findings may not necessarily translate to other similar contexts not implementing the same intervention. As such, one may be unable to predict the effects of similar inputs in different contexts. However, the findings of this study indicate a need for further

research to replicate this type of evaluation in other low-income countries and build an evidence base.

4.4 Areas for Future Research

This study also highlights several exciting areas for future research. These areas include the research into the impact of overage and underage students on EGRA assessment performance within this cohort of students. Additionally, an exciting area of exploration would be to investigate the appropriateness of Rasch analysis for EGRA data – particularly during the development and piloting stages of the assessment design. Finally, controlling for more social-contextual variables within the data offers further learning opportunities.

5 Conclusions

The present study reviewed the assessment of Grade 2 students' reading abilities in Sierra Leone, particularly in relation to social contextual factors. Additionally, the assessment's quality was assessed using a Rasch methodology. The findings above suggest that, while the assessment could be improved, the data that students have exceptionally low reading abilities can be trusted. Additionally, it suggests that reading with a parent may have a particularly large impact on students' performance. Reviewing the descriptive statistics, the quality of the assessment, and critical social-contextual factors help explore student reading outcomes in Sierra Leone. Through this examination, students' literacy outcomes were low—only 5.58 percent of class 2 pupils achieved the reading comprehension benchmark. Pupils performed best on the alphabet naming subtask, but more than half of all pupils struggled with the other subtasks. In general, boys outperformed girls across the subtasks, which indicates a gender-based performance gap. Importantly, low performance on the listening comprehension sub-task demonstrates that students cannot comprehend a story told to them at a level deemed appropriate, showing very

low English abilities that are foundational for reading acquisition. Student performance on the listening comprehension subtask may suggest that students have a limited ability to understand spoken English, which research shows are likely to impact their overall performance on the assessment.

References

Abadzi, H., & Centanni, T. (2020). Why fast and effortless reading is indispensable for comprehension. *Comparative Education Review*, *64*(2), 299–308. https://doi.org/10.1086/708300

Allington, R. L. (1999). What I've learned about effective reading instruction: From a decade of studying exemplary elementary classroom teachers. *Phi Delta Kappan*, *83*(10), 740–747. https://doi.org/10.1177/003172170208301007

Anderson, R. C., Scott, H., & Wilkinson, I. (1985). *Becoming a nation of readers*. 7–18.

Arapa, B., Sánchez, E., Hurtado-Mazeyra, A., & Sánchez, A. (2021). The relationship between access to pre-school education and the development of social-emotional competencies: Longitudinal evidence from Peru. *International Journal of Educational Development, 87*, article 102482

Baker, L., & Wigfield, A. (1999). Dimensions of children's motivation for reading and their relations to reading activity and reading achievement. *Reading Research Quarterly*, *34*(4), 452–477.

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*(January), 308–314. https://doi.org/10.1016/j.ijedudev.2014.10.002

Bennett, K. K., Weigel, D. Cao, F., Wang, Z., Sussman, B. L., Yan, X., Spray, G. J., & Martin, S. S. (2002). Children's acquisition of early literacy skills: Examining family contributions. *Early Childhood Research Quarterly*, *17*(3), 295–317. http://doi.org/10.1016/S0885-2006(02)00166-7

Benson, C. (2002). Real and potential benefits of bilingual programmes in developing countries,

International Journal of Bilingual Education and Bilingualism, 5(6), pp. 303-317.

Bracken, S. S., & Fischel, J. E. (2008). Family reading behavior and early literacy skills in preschool children from low-income backgrounds. *Early Education and Development*, *19*(1), 45–67. http://doi.org/10.1080/10409280701838835

Britto, P., & Brooks-Gunn, J. (2001). Beyond shared book reading: Dimensions of home literacy and low-income African American preschooler's skills. In P. Britto & J. Brooks-Gunn (Eds.), *The role of family literacy environments in promoting young children's emerging literacy skills* (pp. 73–89). Jossey-Bass.

Bronfenbrenner, U. (Ed.). (2005). *Making human beings human: Bioecological perspectives on human development*. SAGE Publications.

Bronfenbrenner, U., & Morris, P. A. (2006). The bioecological model of human development. In R. M. Lerner & W. Damon (Eds.), *Handbook of child psychology: Vol. 1, Theoretical models of human development* (6th ed., pp. 793–828). John Wiley & Sons.

Bus, A. G., van IJzendoorn, M. H., & Pellegrini, A. D. (1995). Joint book reading makes for success in learning to read: A meta-analysis on intergenerational transmission of literacy. *Review of Educational Research*, *65*(1), 1–21. http://doi.org/10.2307/1170476

Burgess, S. R., Hecht, S. A., & Lonigan, C. J. (2002). Relations of the home literacy& Rios, V. (2019). L1 environment (HLE) to the development of reading-related abilities: A one-year longitudinal study. *Reading Research Quarterly*, *37*(4), 408–426. http://doi.org/10.1598/RRQ.37.4.4

J., & Rios, V. (2019). L1 reading experience influences L2 Lexical learning: Spanish learning in Chinese speakers and English speakers. *Neuroscience*, *416*, 255–267.

Catts, H. W., & Kamhi, A. G. (Eds.). (2005). *Language and reading disabilities* (2nd ed.). Pearson.

Chen, J., Zhang, Y., & Hu, J. (2021). Synergistic effects of instruction and affect factors on high- and low-ability disparities in elementary students' reading literacy. *Reading and Writing: An Interdisciplinary Journal, 34(1)*, 199–230. https://doi.org/10.1007/s11145-020-10070-0

Christenson, S. L., & Thurlow, M. L. (2004). School dropouts: Prevention considerations, interventions, and challenges. *Current Directions in Psychological Science*, *13*(1), 36–39. https://doi.org/10.1111/j.0963-7214.2004.01301010.x

Clegg, J. (2005) – Moving towards bilingual education in Africa (Presentation). 7th International Language Development Conference, Addis Ababa, 2015. [Online]

Clegg, J., & Simpson, J. (2016). Improving the effectiveness of English as a medium of instruction in sub-Saharan Africa. Comparative Education, 52(3), pp. 359-374.

Cline, K. D., & Edwards, C. P. (2013). The instructional and emotional quality of parent–child book reading and early head start children's learning outcomes. *Early Education and Development*, *24*(8), 1214–1231. http://doi.org/10.1080/10409289.2012.697431

Coles, G. (2000). *Misreading reading: The bad science that hurts children.* Heinemann Educational Books.

Creemers, B., & Kyriakides, L. (2010). School factors explaining achievement on cognitive and affective outcomes: Establishing a dynamic model of educational effectiveness. *Scandinavian Journal of Educational Research*, *54*, 263–294. https://doi.org/10.1080/00313831003764529

Cunningham, A. E., Perry, K. E., Stanovich, K. E., & Share, D. L. (2002). Orthographic learning during reading: Examining the role of self-teaching. *Journal of Experimental Child Psychology*, *82*(3), 185–199. https://doi.org/10.1016/S0022-0965(02)00008-5

DeBaryshe, B. D., Binder, J. C., & Buell, M. J. (2000). Mothers' implicit theories of early literacy instruction: Implications for children's reading and writing. *Early Child Development and Care*, *160*(1), 119–131. http://doi.org/10.1080/0030443001600111

de Jong, P. F., & van der Leij, A. (2002). Effects of phonological abilities and linguistic comprehension on the development of reading. Scientific Studies of Reading, 6(1), 51–77. http://doi.org/10.1207/S1532799XSSR0601_03

Deming, David. 2009. "Early Childhood Intervention and Life-Cycle Skill Development: Evidence from Head Start." American Economic Journal: Applied Economics, 1 (3): 111-34.

Dodici, B. J., Draper, D. C., & Peterson, C. A. (2003). Early parent–child interactions and early literacy development. *Topics in Early Childhood Special Education*, *23*(3), 124–136. http://doi.org/10.1177/02711214030230030301

Dowd, A. J. M. Y. J. O., Bartlett, L., Khamis-Dakwar, R., & Froud, K. (2020). Measure fluency, vocabulary, and comprehension in international reading efforts. *Comparative Education Review*, *64*(2), 309–315. https://doi.org/10.1086/708301

Duncan, G. J., Dowsett, C. J., Claessens, A., Magnuson, K., Huston, A. C., Klebanov, P., Pagani, L. S., Feinstein, L., Engel, M., Brooks-Gunn, J., Sexton, H., Duckworth, K., & Japel, C. (2007). School readiness and later achievement. *Developmental psychology*, *43*(6), 1428–1446.

Earle, A., Milovantseva, N. & Heymann, J. Is free pre-primary education associated with increased primary school completion? A global study. ICEP 12, 13 (2018). https://doi.org/10.1186/s40723-018-0054-1

Elango, S., García, J. L., Heckman, J. J., & Hojman, A. (2015). Early childhood education (No. w21766). National Bureau of Economic Research.

Eriksson, M., Ghazinour, M., & Hammarström, A. (2018). Different uses of Bronfenbrenner's ecological theory in public mental health research: What is their value for guiding public mental health policy and practice? *Social Theory & Health*, 16, 414–433. https://doi.org/10.1057/s41285-018-0065-6

Francis, D. J., Shaywitz, S. E., Stuebing, K. K., Shaywitz, B. A., & Fletcher, J. M. (1996). Developmental lag versus deficit models of reading disability: A longitudinal, individual growth curves analysis. *Journal of Educational Psychology*, *88*(1), 3–17. https://doi.org/10.1037/0022-0663.88.1.3

Foster, M. A., Lambert, R., Abbott-Shim, M., McCarty, F., & Franze, S. (2005). A model of home learning environment and social risk factors in relation to children's emergent literacy and social outcomes. *Early Childhood Research Quarterly, 20*(1), 13–36. http://doi.org/10.1016/j.ecresq.2005.01.006

Garan, E. M. (2002). *Resisting reading mandates: How to triumph with the truth.* ERIC.

Gates, A. I. (1949). *The Improvement of Reading* (3rd Edition) Macmillan Co.

Gough, P. B., & Tunmer, W. E. (1986). Decoding, reading and reading disability. *Remedial and Special Education*, *7*(1), 6–10. http://dx.doi.org/10.1177/074193258600700104

Gove, A. (2017). What We Are Learning About Early Education sub-Saharan Africa. In *Journal of Research on Educational Effectiveness* (Vol. 10, Issue 3, pp. 530–534). Routledge. https://doi.org/10.1080/19345747.2017.1335091

Gove, A., & Cvelich, P. (2011). *Early Reading: Igniting Education for All. A report by the Early Grade Learning Community of Practice. Revised Edition.*

Gove, A., & Wetterberg, A. (Eds.). (2011). *The early grade reading assessment: Applications and interventions*. TRI International. https://doi.org/10.3768/rtipress.2011.bk.0007.1109

The Education Act, (2004) (testimony of Government of Sierra Leone).

Government of Sierra Leone. (2018). *Education Sector Plan 2018-2020*.

Graham, B. E., & van Ginkel, A. J. (2014). Assessing early grade reading: the value and limits of 'words per minute.' *Language, Culture and Curriculum*, *27*(3), 244–259.

Grantham-McGregor, S., Cheung, Y. B., Cueto, S., Glewwe, P., Richter, L., Strupp, B., & International Child Development Steering Group (2007). Developmental potential in the first 5 years for children in developing countries. *Lancet (London, England)*, *369*(9555), 60–70. https://doi.org/10.1016/S0140-6736(07)60032-4

Halliday, K. E., Karanja, P., Turner, E. L., Okello, G., Njagi, K., Dubeck, M. M., Allen, E., Jukes, M. C. H., & Brooker, S. J. (2012). Plasmodium falciparum, anaemia and cognitive and educational performance among school children in an area of moderate malaria transmission: baseline results of a cluster randomized trial on the coast of Kenya. *Tropical Medicine & International Health*, *17*(5), 532–549.

Hambleton, R. K., Swaminathan, H., & Rogers, H. J. (1991). *Fundamentals of item response theory* (Vol. 2). SAGE Publishing.

Hanushek, E. A., & Woessmann, L. (2012). Do better schools lead to more growth? Cognitive skills, economic outcomes, and causation. *Journal of Economic Growth*, *17*(4), 267–321. https://doi.org/10.1007/s10887-012-9081-x

Hernandez, D. J. (2011). Double jeopardy: How third grade reading skills and poverty influence high school graduation. *Annie E. Casey Foundation*, 1–15.

Ho, S., Chan, D., Chung, K., Tsang, S., Lee, S., & Cheng, W. Y. R. (2007). The Hong Kong Test of Specific Learning Difficulties in Reading and Writing for Primary School Students (2nd ed.).

Hoffman, J. V. (2012). Standpoints: Why EGRA-a Clone of DIBELS-Will Fail to Improve Literacy in Africa. In *Source: Research in the Teaching of English* (Vol. 46, Issue 4).

Hood, M., Conlon, E., & Andrews, G. (2008). Preschool home literacy practices and children's literacy development: A longitudinal analysis. *Journal of Educational Psychology*, *100*(2), 252–271. http://doi.org/10.1037/0022-0663.100.2.252

Hoover, W. A., & Gough, P. B. (1990). The simple view of reading. *Reading and Writing*, *2*(2), 127–160. https://doi.org/10.1007/BF00401799

Huey, E. B. (1968). *The psychology and pedagogy of reading*. MIT Press. (Original work published 1908)

Kamii, C., & Manning, M. (2005). Dynamic Indicators of Basic Early Literacy Skills (DIBELS): A tool for evaluating student learning? *Journal of Research in Childhood Education*, *20*(2), 75–90. http://doi.org/10.1080/02568540509594553

Krafft, C. (2015) Increasing educational attainment in Egypt: The impact of early childhood care and education. *Economics of Education Review, 46*: 127–143.

Krashen, S. (2002). Defending whole language: The limits of phonics instruction and the efficacy of whole language instruction. *Reading Improvement*, *39*(1), 32–42.

Leonard, J. (2011). Using Bronfenbrenner's ecological theory to understand community partnerships: A historical case study of one urban high school. Urban Education, 46, 987–1010. https://doi.org/10.1177/0042085911400337

Luppescu, S., & Ehrlich, S. (2012). *The new and improved Rasch measurement model primer*. UChicago Consortium on School Research. https://consortium.uchicago.edu/sites/default/files/2019-05/Overview%20of%20Analyses.pdf

MacDonald, A., & Murphy, S. (2019). Mathematics education for children under four years of age: A systematic review of the literature. *Early Years* 1–18.

Manji, S. (2018). Getting Early Grade Reading Right: A case for investing in quality early childhood education programs. *Paper prepared for USAID Building Evidence and Supporting Innovation to Improve Primary Grade Reading Assistance* for the Office of Education (E3/ED), University Research Co., LLC.

McCoy, D. Zulkowski, S., Yoshikawa, H. & Günther, F. (2017) Early Childhood Care and Education and School Readiness in Zambia*, Journal of Research on Educational Effectiveness, 10:3*, 482-506, DOI: 10.1080/19345747.2016.1250850

McCoy, D. C., Yoshikawa, H., Ziol-Guest, K. M., Duncan, G. J., Schindler, H. S., Magnuson, K., Yang, R., Koepp, A., & Shonkoff, J. P. (2017). Impacts of early childhood education on medium-and long-term educational outcomes. *Educational Researcher, 46*(8), 474–487.

Ministry of Health and Sanitation, S. L. (2013). *Sierra Leone Demographic and Health Survey*.

Mohamed, A., & Carthage, T. (2013). The role of Arabic orthographic literacy in the phonological awareness of Tunisian children. *International Journal on New Trends in Education and Their Implications*.

Mohamed, Kamanda (2002) Mother Tongue Education and Transitional Literacy in Sierra Leone: Prospects and Challenges in the 21st Century, *Language and Education, 16:3,* 195-211, DOI: 10.1080/09500780208666828

Mol, S. E., Bus, A. G., de Jong, M. T., & Smeets, D. J. H. (2008). Added value of dialogic parent–child book readings: A meta-analysis. *Early Education and Development*, *19*(1), 7–26. http://doi.org/10.1080/10409280701838603

Moore, A. M., Gove, A., & Tietjen, K. (2017). Great expectations: A framework for assessing and understanding key factors affecting student learning of foundational reading skills. *New Directions for Child and Adolescent Development*, *2017*(155), 13–30. https://doi.org/10.1002/cad.20192

Nag, S., Vagh, SB, Dulay, K, & Snowling, M. (2018). Home language, school language and children's literacy attainments: A systematic review of evidence from low- and middle-income countries. *Review of Education, 7*, 91-150.

Nonoyama-Tarumi, Y., & Bredenberg, K. Impact of school readiness program interventions on children's learning in Cambodia*. International Journal of Educational Development*, 29, no. 1 (2009): 39–45.

Protopapas, A., Simos, P. G., Sideridis, G. D., & Mouzaki, A. (2012). The components of the simple view of reading: A confirmatory factor analysis. *Reading Psychology*, *33*(3), 217–240. http://dx.doi.org/10.1080/02702711.2010.507626

Rao N, Sun J, Wong JMS, Weekes B, Ip P, Shaeffer S, Young M, Bray M, Chen E, Lee D (2014) Early childhood development and cognitive development in developing countries: A rigorous literature review. Department for International Development.

Roberts, J., Jurgens, J., & Burchinal, M. (2005). The role of home literacy practices in preschool children's language and emergent literacy skills. *Journal of Speech, Language, and Hearing Research*, *48*(2), 345–359. http://dx.doi.org/10.1044/1092-4388(2005/024)

Rush, K. L. (1999). Caregiver–child interactions and early literacy development of preschool children from low-income environments. *Topics in Early Childhood Special Education*, *19*(1), 3–14. http://doi.org/10.1177/027112149901900101

Samuels, S. J. (2007). The DIBELS tests: Is speed of barking at print what we mean by reading fluency? *Reading Research Quarterly*, *42*, 563–566.

Shelton, N. R., Altwerger, B., & Jordan, N. (2009). Does DIBELS put reading first? *Literacy Research and Instruction*, *48*(2), 137–148. http://doi.org/10.1080/19388070802226311

Skibbe, L. E., Justice, L. M., Zucker, T. A., & McGinty, A. S. (2008). Relations among maternal literacy beliefs, home literacy practices, and the emergent literacy skills of preschoolers with specific language impairment. *Early Education and Development*, *19*(1), 68–88.http://doi.org/10.1080/10409280701839015

Sonnenschein, S., & Munsterman, K. (2002). The influence of home-based reading interactions on 5-year-olds' reading motivations and early literacy development. *Early Childhood Research Quarterly*, *17*(3), 318–337. http://doi.org/10.1016/S0885-2006(02)00167-9

State House Media and Communications Unit. (2018). *President Bio Launches Free Education, Calls on Parents and Teachers to Support the Initiative – Sierra Leone State House – The*

Republic of Sierra Leone State House. https://statehouse.gov.sl/president-bio-launches-free-education-calls-on-parents-and-teachers-to-support-the-initiative/

Tabouret-Keller Andrée. (1997). Vernacular literacy : a re-evaluation : by members of the international group for the study of language standardization and the vernacularization of literacy (iglsvl). Clarendon Press.

Tobia, V., & Bonifacci, P. (2015). The simple view of reading in a transparent orthography: The stronger role of oral comprehension. *Reading and Writing*, *28*, 939–957. http://doi.org/10.1007/s11145-015-9556-1

Torrente, C., Aber, J. L., Starkey, L., Johnston, B., Shivshanker, A., Weisenhorn, N., Annan, J., Seidman, E., Wolf, S., & Tubbs Dolan, C. (2019). Improving primary education in the Democratic Republic of the Congo: End-line results of a cluster-randomized wait-list controlled trial of learning in a healing classroom. *Journal of Research on Educational Effectiveness*, *12*(3), 413–447. https://doi.org/10.1080/19345747.2018.1561963

Trudell, B., & Schroeder, L. (2007a). Reading methodologies for African languages: Avoiding linguistic and pedagogical imperialism. *Language, Culture and Curriculum*, *20*(3), 165–180.

Trudell, B., & Schroeder, L. (2007b). Reading methodologies for African languages: Avoiding linguistic and pedagogical imperialism. *Language, Culture and Curriculum*, *20*(3), 165–180.

Tudge, J. R. H., Mokrova, I., Hatfield, B. E., & Karnik, R. B. (2009). Uses and misuses of Bronfenbrenner's bioecological theory of human development. *Journal of Family Theory & Review*, *1*(4), 198–210. https://doi.org/10.1111/j.1756-2589.2009.00026.x

Thulla, Philip Foday Yamba & Leone, Sierra & Moriba, Samba & Adom, Dickson & Mensah-Gborie, Madiana. (2022). The Rate of Reading Poverty After the COVID- 19 Pandemic School Shutdown and Specific Intervention Strategies for Lower Primary School Pupils in the Southern Province and Western Area of Sierra Leone. Journal of Language Teaching and Research. 13. 689-696. 10.17507/jltr.1304.01.UNESCO. (2018). *Quick Guide No. 2: Making the Case for a Learning Assessment*.

UNESCO (1953) 'The Use of Vernacular Languages in Education', Monographs on Foundations of Education, No. 8. Paris: UNESCO [Online] http://unesdoc.unesco.org/images/0000/000028/002897eb.pdf

UNESCO Global Monitoring Report (GMR) (2015) Education for All 2000-2015: achievements and challenges; EFA global monitoring report, 2015. Paris: UNESCO.

UNESCO. (2019a). Migration, Displacement and Education. In *Global Education Monitoring Report*. https://unesdoc.unesco.org/ark:/48223/pf0000366946/PDF/366946eng.pdf.multi

UNESCO. (2019b). *Sierra Leone Country Statistics*. UNESCO Institute of Statistics. http://uis.unesco.org/country/SL#slideoutmenu

United Nations General Assembly (2015). Transforming our world: The 2030 Agenda for Sustainable Development, 21 October 2015, http://www.refworld.org/docid/57b6e3e44.html [accessed 1 October 2018].

Vandecandelaere, M., Vanlaar, G., Goos, M., De Fraine, B., & Van Damme, J. (2013). Effects of grade retention on growth in mathematics in the third year of kindergarten: A propensity score stratification analysis. *Leuven:* Steunpunten voor Beleidsrelevant Onderzoek.

van Huizen, Thomas & Plantenga, Janneke, (2018), Do children benefit from universal early childhood education and care? A meta-analysis of evidence from natural

experiments, *Economics of Education Review*, 66, issue C, p. 206-222, https://EconPapers.repec.org/RePEc:eee:ecoedu:v:66:y:2018:i:c:p:206-222.

Wagner, D. A., Babson, A., & Murphy, K. M. (2011). How much is learning measurement worth? Assessment costs in low-income countries. . *Current Issues in Comparative Education*, *14*,(1), 3—23.

Wagner, D. A., Lockheed, M., Mullis, I., Martin, M. O., Kanjee, A., Gove, A., & Dowd, A. J. (2012). The debate on learning assessments in developing countries. *Compare: A Journal of Comparative and International Education*, *42*(3), 509–545. https://doi.org/10.1080/03057925.2012.670480

Walberg, H. J. (1984). Improving the productivity of America's schools. Educational Leadership, 41, 19–27.

Walter, S., & Benson, C. (2012). Language policy and medium of instruction in formal education. In B. Spolsky (Ed.), The Cambridge Handbook of Language Policy (Cambridge Handbooks in Language and Linguistics, pp. 278-300). Cambridge: Cambridge University Press. doi:10.1017/CBO9780511979026.017

Weigel, D. J., Martin, S. S., & Bennett, K. K. (2006). Contributions of the home literacy environment to preschool-aged children's emerging literacy and language skills. *Early Child Development and Care*, *176*(3-4), 357–378. http://doi.org/10.1080/03004430500063747

Weisleder A, Fernald A. Talking to children matters: early language experience strengthens processing and builds vocabulary. Psychol Sci. 2013 Nov 1;24(11):2143-52. doi: 10.1177/0956797613488145. Epub 2013 Sep 10. PMID: 24022649; PMCID: PMC5510534.

Wright, B. D., & Stone, M. H. (1979). Best Test Design: Rasch Measurement. Chicago, IL: Mesa Press.

Wright, B. D. (1996) Comparing Rasch measurement and factor analysis, Structural Equation Modeling: A Multidisciplinary Journal, 3:1, 3-24, DOI: 10.1080/10705519609540026

Yoshikawa, H. et al. (2013) Investing in Our Future: The Evidence Base on Preschool Education. *Society for Research in Child Development.*

Zajacova, A., & Lawrence, E. M. (2018). The relationship education and health: reducing disparities through a contextual approach. *Annual Review of Public Health*, *39*, 273–289. https://doi.org/10.1146/ANNUREV-PUBLHEALTH-031816-044628