

Paper prepared for the
Global Education Monitoring Report

2021 GEM Report Fellowship

Language considerations for the global monitoring agenda: Insights from a large-scale bilingual assessment of literacy from the Democratic Republic of the Congo

This paper was developed as part of the GEM Report Fellowship Programme. It has not been edited by the team. The views and opinions expressed in this paper are those of the author(s) and should not be attributed to the Global Education Monitoring Report or to UNESCO. The papers can be cited with the following reference: “Paper commissioned as part of the GEM Report Fellowship Programme in 2021”. For further information, please contact gemreport@unesco.org.

FERNANDA GÁNDARA

2022

ABSTRACT

Assessments of literacy used in global monitoring are typically monolingual, even in contexts that have adopted bilingual education. Preliminary research shows that bilingual assessments – or assessments that aim to capture the full linguistic repertoire of bilingual students – may be fairer and more appropriate than monolingual assessments. However, the world is far from using bilingual assessments in higher stakes circumstances. There is not a clear path for transitioning from monolingual to bilingual assessment frameworks. More evidence is needed to consider bilingual assessments for evaluation purposes such as global monitoring. This paper uses data from a bilingual assessment to evaluate different representations of bilingual literacy in a linguistically diverse context like the Democratic Republic of the Congo. In addition, this study examines whether bilingual assessments of literacy provide equivalent information for students who come from different linguistic backgrounds, across multiple provinces. The results show that complex multidimensional conceptualizations of literacy may best represent bilinguals' proficiency. Such complex representations can only be captured via bilingual assessments. However, this study suggests that bilingual assessments are not inherently equivalent across subgroups of students. The outcomes vary by provincial and language groups, raising doubts as to whether bilingual assessments could serve high-stake large-scale uses. Policy recommendations are provided, considering the current challenges that bilingual education faces in contexts like the DRC. Most notably, countries should consider using bilingual assessments as a part of their efforts to strengthen bilingual education programs.

TABLE OF CONTENT

| | |
|--|-----------|
| Abstract | 1 |
| Table of Content | 2 |
| TABLES | 4 |
| APPENDIX TABLES | 4 |
| FIGURES | 5 |
| 1. Introduction | 6 |
| 1.1. Purpose | 7 |
| 2. Literature Review | 8 |
| 2.1. Bilingual education in Sub-Saharan Africa | 8 |
| 2.2. Considerations for assessing literacy in Sub-Saharan Africa | 10 |
| 2.3. The potential of bilingual assessments | 10 |
| 3. Context | 12 |
| 3.1. Education in the DRC | 12 |
| 3.2. Language-in-education policies in the DRC | 13 |
| 3.3. Assessments in the DRC | 14 |
| 3.4. Literacy outcomes in the DRC | 15 |
| 3.4.1. MICS | 15 |
| 3.4.2. EGRA | 17 |
| 3.4.3. ACCELERE! Quarterly Assessments – Grade 2 | 18 |
| 3.4.4. PASEC | 19 |
| 3.4.5. Proficiency across assessments | 19 |
| 4. Data and methodology | 21 |
| 4.1. Data | 21 |
| 4.1. Analytical Framework | 23 |
| 4.1.1. First research question – Challenges faced by bilingual education | 23 |
| 4.1.2. Second research question – Appropriateness of bilingual assessments | 24 |
| 4.1.3. Third research question – Equivalence of bilingual assessments | 26 |
| 5. Literature review findings | 28 |
| 5.1.1. Sociocultural challenges | 28 |
| 5.1.2. Policy design challenges | 31 |
| 5.1.3. Resource challenges | 31 |
| 5.1.4. Ideology and conceptual challenges | 32 |
| 5.1.5. Linguistic challenges | 34 |
| 5.1.6. Historical challenges | 35 |
| 5.1.7. Implementation challenges | 36 |

| | |
|--|-----------|
| 6. Results | 37 |
| 6.1. Research question 2: Representation of bilingual data | 37 |
| 6.2. Research question 3: Equivalence of bilingual assessments | 43 |
| 7. Conclusions | 45 |
| Policy Recommendations | 49 |
| Acknowledgements | 51 |
| References | 52 |
| Appendices | 60 |
| Appendix 1: SEM Results | 60 |

TABLES

| | |
|---|----|
| Table 1 Percentage demonstrating foundational knowledge in reading, MICS, Children Aged 7-14..... | 16 |
| Table 2 Percentage demonstrating foundational knowledge in reading, MICS, Grade 2 and Grade 3 students..... | 17 |
| Table 3 EGRA 2018 Results – by Provincial Group..... | 18 |
| Table 4 Quarterly Assessment Results – Oral Reading, Grade 2 - 2018 | 18 |
| Table 5 Quarterly Assessment Results – Reading Comprehension, Grade 2 - 2018 | 19 |
| Table 6 Characteristics of the Dataset..... | 21 |
| Table 7 Tasks in the ACCEERE! Quarterly Assessments..... | 22 |
| Table 8 Models and Interpretation..... | 25 |
| Table 9 Fit Indices and Interpretation..... | 26 |
| Table 10 Performance by Task and Provincial Group (Percent Correct) | 39 |
| Table 11 χ^2 test for bidimensional and tridimensional models..... | 40 |
| Table 12 P-values for improvement in fit if restriction lifted, by parameter | 44 |
| Table 13 Additional Recommendations Emerging from the Analysis of Bilingual Data | 50 |

APPENDIX TABLES

| | |
|---|----|
| Appendix Table 1 Model Comparison by Province..... | 60 |
| Appendix Table 2 Fit Analysis Ciluba | 60 |
| Appendix Table 3 Fit Analysis Lingala | 60 |
| Appendix Table 4 Fit Analysis Kiswahili | 61 |
| Appendix Table 5 Multigroup Analysis, Linguistic Diversity at Home, Ciluba..... | 61 |
| Appendix Table 6 Multigroup Analysis, Linguistic Diversity at Home, Lingala | 61 |
| Appendix Table 7 Multigroup Analysis, Linguistic Diversity at Home, Kiswahili | 61 |
| Appendix Table 8 Multigroup Analysis, Type of Language, Ciluba | 62 |

| | |
|---|----|
| Appendix Table 9 Multigroup Analysis, Type of Language, Lingala..... | 62 |
| Appendix Table 10 Multigroup Analysis, Type of Language, Kiswahili | 62 |
| Appendix Table 11 Invariance Analysis, Fit on Focal Group, Linguistic Diversity | 62 |
| Appendix Table 12 Invariance Analysis, Fit on Focal Group, Minority Languages | 62 |

FIGURES

| | |
|---|----|
| Figure 1 National Level EGRA Assessments Administered in the DRC, by Grade Level..... | 17 |
| Figure 2 Percentage of Students Meeting Expectation using Different Assessments and Proficiency Criteria, Grade 2 Students..... | 20 |
| Figure 3 Final Selection of Articles by Geographic Focus | 24 |
| Figure 4 Proportion of Students by Number of Languages Spoken at Home..... | 27 |
| Figure 5 Proportion of Students by Minority Language Spoken at Home or Not..... | 27 |
| Figure 6. Predominance of Challenges in the Reviewed Documents | 28 |
| Figure 7 Sample Items..... | 37 |
| Figure 8 RMSEA and CFI for Models | 39 |
| Figure 9 Percentage of Misfit for Different Models..... | 40 |
| Figure 10 Best Model, Ciluba | 41 |
| Figure 11 Best Model, Lingala | 42 |
| Figure 12 Best Model, Kiswahili | 43 |
| Figure 13 Invariance Analyses on the Best Performing Model..... | 44 |

1.Introduction

The United Nations' Sustainable Development Goals (SDGs) and their associated framework of indicators are rapidly shifting the priorities in the education sector worldwide. The SDG monitoring framework is multi-tiered, and demands national, regional, thematic, and global levels of reporting (UIS-UNESCO, 2017). The most prominent global indicator for SDG 4 is indicator 4.1.1 which monitors the percentage of students meeting proficiency standards in reading and mathematics (United Nations, 2015). This indicator is of paramount importance, since too many children around the world do not achieve minimum standards of proficiency (World Bank, 2018). However, indicator 4.1.1 poses comparability challenges, as measuring proficiency is not equally resolved or approached with the same level of rigor across countries. A student could be considered proficient in Country A and not proficient in Country B, if the standards or benchmarks for proficiency focus on different skills and/or capture varying levels of difficulty. Consequently, the international community has developed methodological guidance for countries to use as they report against indicator 4.1.1.

Efforts to improve monitoring of SDGs and particularly, indicator 4.1.1 are significant. However, there are threats to the validity of SDG4 indicators that rest on the educational assessments themselves. Often overlooked is the language of those assessments. Large-scale standardized assessments, such as those that would feed SDG4 indicators, are mostly monolingual in focus (Shohamy, *Imagined multilingual schools: How come we don't deliver*, 2006). Monolingual assessments used in bilingual contexts are questionable on ethical grounds, as they neglect the multiple languages of students and could serve ideologies that seek assimilation through language (Shohamy, 2007; Shohamy, 2011). Monolingual assessments used in bilingual¹ contexts may fail to meet their most fundamental goal, which is to provide information on the proficiency of students. Assessing students who are not fluent in the language of the assessment makes it difficult to disentangle whether the observed performance is due to construct relevant variance or to language barriers (Abedi, 2004). Moreover, using monolingual assessments with students who are functional in the language used may mask the fact that bilinguals' linguistic repertoire cannot be fully captured in any one language. As Sireci (2020) recently pointed out, standardization as a principle may be too rigid to properly accommodate the heterogeneity of students, which strongly applies to the concept of language as well. This is not to say that the SDG framework neglects the relevance of linguistic heterogeneity. Indeed, the SDG framework acknowledges the importance of language, particularly through indicator 4.5.2 which measures the percentage of students whose home language is the language of instruction (Technical Cooperation Group (TCG), 2018). Equity indicators such as SDG 4.5.2 are relevant, as they recognize that students who do not speak the language of a test at home underperform their peers (GEM Report, 2016; Howie & Chamberlain, 2017). By using indicator 4.5.2 alongside indicator 4.1.1, practitioners and policy makers can grasp the extent to which the latter is an equitable measure of proficiency. However, the situation is far more complex than dichotomizing students into

¹ In this paper, bilingual and multilingual are synonyms

“speakers of the language or not.” Students who are exposed to multiple languages at home may be considered speakers of a given language and at the same time, be at disadvantage in relation to peers who only speak that language at home. In addition, speakers of multiple languages may possess more linguistic resources than peers yet be less proficient in each of their languages when compared to monolingual students. Unless bilingual students are assessed on their full repertoires and unique set of skills, they will remain at disadvantage.

1.1.Purpose

Comparative monitoring of student performance around the world is not a straightforward task. Tracking global indicators such as SDG 4.1.1, requires a common understanding of what proficiency ought to be across the world. In addition, countries necessitate data from national level assessments of literacy and numeracy to report into these indicators, which are often not in place. As global monitoring expands and more countries try to adopt large-scale assessments, it is paramount to revisit unsettled validity issues that could as well impact global monitoring practice, such as linguistic choices in assessments. High-stakes monolingual large-scale assessments can be detrimental for bilingual students and for bilingual education, more generally. For instance, using monolingual assessments for monitoring could result in the prioritization of teaching and learning in the language of the assessment (i.e., “teaching to the test”), regardless of country specific language-in-education policies. However, there is neither enough research comparing monolingual and bilingual assessments, nor enough guidance on how to transition from a monolingual to a bilingual assessment framework.

The purpose of this study is to enrich the conversation around measuring literacy and around global monitoring from the perspective of language used in assessments. This study frames the problem of linguistic choices within the broader context of bilingual education. Doing so sheds light on the true potential, barriers, and relevance of linguistic choices. The study uses data from the Democratic Republic of the Congo (DRC). The DRC is an appropriate case study for examining bilingual assessments, as the country has a rich ethnolinguistic diversity and has adopted a bilingual form of education that recognizes and promotes several languages in early grades. In addition, the DRC is a low-income country where resources are often lacking, and education does not reach all subgroups of the population. Thus, the DRC is a well-suited context to explore the challenges and opportunities of bilingual education and assessments. This study is guided by the following research questions:

1. What are some of the critical challenges that affect bilingual education programs in Sub-Saharan Africa, including the Democratic Republic of the Congo?
2. How should literacy be represented for bilingual students, based on data from a large-scale bilingual assessment?
3. Are large-scale bilingual assessments necessarily equivalent for students from different linguistic backgrounds?

2. Literature Review

2.1. Bilingual education in Sub-Saharan Africa

Bilingual² education refers to education in more than one language. Baker (2011) proposes three categories of education programs. First, monolingual forms of education, which promote and value proficiency in one language; second, weak forms of bilingual education, which encourage bilingual practices while one language is valued more than the other(s) and the outcome is rarely bilingualism; and third, strong bilingual programs, where the aim is for students to become bilingual (Baker, 2011). Monolingual programs are restrictive and view other languages – typically minority languages – as problems (Mwaniki, Arias, & Wiley, 2017). Examples of monolingual programs are *submersion* programs, in which majority languages are imposed on the whole population of students. Weak bilingual programs can be thought of as tolerant, under which there is some room for a minority language that is not effectively promoted (Mwaniki, Arias, & Wiley, 2017). An example of a weak bilingual program is a *separatist* program where students can choose a minority language to be used in instruction. And strong bilingual programs can be thought of as those truly promoting bilingualism (Mwaniki, Arias, & Wiley, 2017), in which minority languages are seen as a resource and not as a problem. An example of a strong bilingual program is a *two-way / dual language program* that promotes pluralism, and students are expected to become fully bilingual.

A different framework to think about bilingual programs is the one used by García (2009). Under this framework, bilingual programs can be viewed as *subtractive* programs through which one's first language is lost and a second language is added, resulting in children only speaking the second language. Programs can also be viewed as *additive*, under which a second language is added to the repertoire and maintained through schooling. As explained by García (2009), both subtractive and additive models follow *monoglossic* ideologies – which value each of the languages in relation to monolingual standards. For instance, additive models follow a monolingual view of languages, because students eventually become “double monolinguals.” Additionally, programs can be thought of as *recursive*, where school bilingualism is conceived as part of a broader effort to revitalize a language used by a community in different degrees (e.g., an old indigenous language). In these programs, students move “back and forth along a bilingual continuum” (p.68, García, 2009) as opposed to fully acquiring a second language in school. This model originates from a *heteroglossic* ideology – where multiple language practices coexist in interrelationship and multiple voices are recognized. Last, García (2009) refers to *dynamic* bilingual programs, in which bilingualism is considered a dynamic phenomenon, recognizing that students have linguistic repertoires with elements from different languages, that they use strategically in complex communicative ways. Programs that embrace dynamic bilingualism create opportunities for students to engage in complex multimodal multilingual practices. Under this model, bilingualism is a resource and hybrid cultural experiences are valued (García, 2009).

² In this article, bilingualism and multilingualism mean more than one language.

In Sub-Saharan Africa, most countries use an *early-exit model* of bilingual education, which consists of 1 to 3 years of using an African language as language of instruction (LOI), after which a full transition is made to using an international or foreign language as LOI (Masagazi Masaazi, Ssentanda, & Ngaka, 2018). For instance, Gambia uses a first language (L1) or *area language* (local language) as LOI until grade 3, after which English is used (Masagazi Masaazi, Ssentanda, & Ngaka, 2018). Uganda uses multiple LOIs during grades 1-3 in rural areas, and Kiswahili in urban areas; while in all regions, English becomes the LOI after grade 3 (Masagazi Masaazi, Ssentanda, & Ngaka, 2018). In Kenya, the LOI for lower primary schooling also varies depending on the urbanicity of the region, with either English (urban), Kiswahili (multi-ethnic rural), or indigenous languages (homogeneous rural areas) used until grade 3, after which English becomes the official LOI (Muaka, 2011). South Africa recognizes eleven official languages that can be used as LOI in early grades, with English or Afrikaans becoming the LOI starting at grade 4 (Evans, 2018). The DRC recognizes four National Languages that can be used as LOI until grade 3, after which French, the official language, becomes the LOI (Journal Officiel de la RDC, 2011). And Ghana uses a Ghanaian language (area language) or English in lower primary, and fully switches to English in grade 4 (Nyarko Ansah, 2014).

Not all African countries follow a transitional model. For instance, in Somalia, “mother tongue” education is provided through secondary education (Masagazi Masaazi, Ssentanda, & Ngaka, 2018). In contrast, countries such as Rwanda or Malawi use English – a foreign language – as the LOI since grade 1 (Nishioka & Durrani, 2018; Williams, 2021). And in some other countries such as Tanzania, an African language (Kiswahili) is the official LOI of primary and secondary schooling (Reddick & Dryden-Peterson, 2021). Importantly, these policies do not have direct correspondence with the linguistic heterogeneity of the national populations. In conclusion, despite a predominant model in the region, significant country-level nuances exist. For the most part, education programs in Sub-Saharan Africa are monoglossic in orientation.

Transitional models used in Sub-Saharan Africa, such as the early-exit model, can take multiple forms. After the transition, these models can be considered weak bilingual programs if the African languages continue to be used across grade levels (e.g., *maintenance* programs), or monolingual, if these languages disappear from the curriculum (e.g., *fully transitional*). Using García’s framework, these models can either be subtractive or additive, depending on the role that the African language plays after the transition is made. Often, these models are referred to as *mother tongue early exit* programs, because they are supposedly using *mother tongues* – languages learned by a child at home (Trudell, 2016) – in the earlier years. However, the term mother tongue is contested, particularly in Africa, because the language learned at home may not correspond to the language spoken by one’s mother or the language spoken by the community (Ekkehard, 2018). The term has also been criticized because it is only used for minority languages, carrying currency (Baker & Prys Jones, 1998). Moreover, in linguistically heterogeneous contexts such as most of Sub-Saharan Africa, the language used for schooling may be no one’s home language or first language. As such, the concept of mother tongue education is useful and valid but should be used with caution, as it may provide the false notion that students are learning in a home language, when that is rarely the case (Orwenjo, 2012).

2.2. Considerations for assessing literacy in Sub-Saharan Africa

Literacy is an essential skill for the future wellbeing of students, yet children in low – and middle – income countries are far from achieving the minimum standards of literacy (World Bank, 2018). As a result, donors and international organizations have increased the emphasis on supporting students to acquire the foundational literacy they would need to succeed later in life. The increased emphasis on foundational learning, and the commitments carried by the SDGs, have led to more frequent assessments of literacy administered to young students around the world. The proliferation of literacy assessments carries some challenges for bilingual contexts such as most Sub-Saharan countries.

Literacy can be defined as “the ability to construct meaning from a text” (Hedgcock & Ferris, 2009), underlining reading comprehension. Reading comprehension can be thought of as a skill that is acquired after mastering a series of other skills – a.k.a. essential components – such as letter sound knowledge, and phonemic awareness or decoding, to name a few. However, the relevance and the roles that each of these skills play in learning how to read depend on whether one is learning to read in a first (L1) or second language (L2), and also on the linguistic characteristics of each language. For example, learning how to read in a *transparent language* – i.e., a language where there is clear correspondence between sounds and their written representation – is different than learning how to read in an *opaque language* – where such correspondence is not evident. Decoding in a transparent language like most African languages, involves relying on small units of text (letter-phoneme conversions or syllables), whereas decoding in English means relying on longer strings (Schroeder, 2012). On the other hand, students who learn how to read in a second language often benefit from knowing how to read in their first language (Cummins, 1981). Such hypothesis – a.k.a. Cummins’ *interdependence hypothesis* – is supported by evidence from multiple contexts and languages. For example, a recent study using transparent languages in rural South Africa shows that lower order skills such as phoneme isolation and letter-sound recognition in L1 “transfer” to support the development of L2 (English) phoneme awareness, letter knowledge, word reading and spelling (Schaefer & Kotze, 2019). These considerations are extremely relevant to developing literacy assessments in the context of Sub-Saharan Africa, where students are typically learning how to read in multiple languages of widely distinct characteristics.

2.3. The potential of bilingual assessments

Monolingual assessments may not be appropriate to evaluate the proficiency of bilingual students. Bilingual students – or students who have unique experiences using more than one language (García, 2009) – are rarely assessed on the full range of skills that they master, including metacognitive skills. Students are usually assessed in languages prioritized in each education system, which could be unfair if they are not proficient in those languages. Efforts to improve fairness via mother tongue assessments may fall short for other reasons. For example, in contexts such as Sub-Saharan Africa, it is common to assess students’ literacy using assessments in “mother tongues,” yet these assessments are merely translations of tests in dominant languages, and do not account for the specificities of the language(s) tested or the skills that matter in their acquisition (Schroeder, 2012). Translation is not the same as

sociolinguistic adaptation. Even if properly adapted, bilingual students have an integrated set of skills that is not captured by one or many disconnected monolingual assessments.

At the minimum, assessments of literacy administered to bilingual students should account for their full proficiency. There are two approaches that could be followed (Shohamy, 2011). First, an approach that views each language as a closed and clearly defined set: for example, French and English are two distinct languages and students may know some French, some English and some skills that span both languages. Second, an approach that deconstructs the notion of finite languages and views the different linguistic resources of bilinguals as a unified set. In this article, we focus on the first approach only, as it has been the predominant framework used in international education. Following this approach, research shows that bilingual students might know some things in one language, some things in the other language, and some things in both languages (Páez, 2008), and as a result, monolingual assessments would not be able to fully capture their skills. In addition, there is important variability in the competencies of bilinguals across their common languages (Sanchez et al., 2013). For example, bilinguals may have different levels of oracy and literacy among their multiple languages or may follow different developmental trajectories in their language acquisition (García, 2009). Therefore, bilingual assessments may measure different skills in each language. Importantly, assessments of literacy for students acquiring a second (or third) language, should also examine their skills in L1s. Assessments that do not examine the full linguistic repertoire of bilinguals, or do not recognize multiple developmental paths, may underestimate their proficiency.

Increasingly, practitioners are adopting “double monolingual” practices that examine the proficiency of bilinguals across multiple languages by administering separate monolingual assessments that typically measure the same skills. Double monolingual assessment methods enable practitioners to measure the relative competence of students across languages and evaluate whether their bilingualism is balanced. Ultimately, double monolingual practices support statements such as “students’ literacy in their L1 is higher than their literacy in their L2.” In principle, multiple monolingual assessments could capture bilinguals’ skills across their multiple languages. However, double, or multiple monolingual assessments are not necessarily the same as bilingual assessments, because the latter are oriented towards capturing the complexity of bilinguals’ linguistic repertoire. Bilinguals use language in ways that do not equate to double monolingualism (García, 2009). For instance, bilinguals codeswitch, borrow terms, and make specific types of mistakes. The heterogeneity in language use could be more marked in earlier grades, when students’ languages are emerging in uneven manners. Notably, double monolingual assessments fail to capture the unique cross-linguistic ways in which bilinguals use language, particularly at earlier grades.

Assessments that aim to evaluate the full linguistic repertoire of bilinguals, including their skills across multiple languages, and in an integrated and developmentally appropriate manner – herein referred to as *bilingual assessments* – could address some of the current shortcomings. In literacy evaluation, bilingual assessments that capture skills in several languages but also across languages, via bilingual and metalinguistic items, could provide a

fuller picture of students' skills. As such, bilingual assessments could better support pedagogy within bilingual education systems. In particular, the integrated focus of bilingual assessments could give a more nuanced picture of students' literacy development across languages and provide diagnostic and formative information for educators. Bilingual assessments could better serve monitoring and policy, by providing more comprehensive insight into students' learning. Moreover, evidence suggests that bilingual assessments can be beneficial for bilingual students in terms of performance. For example, Páez (2008) shows that accounting for students' first language skills in addition to their English skills leads to improvements in language tests. Similarly, Sanchez et al. (2013) argue that measuring bilinguals' fuller linguistic repertoire leads to a better understanding of the academic achievements of students. And, in mathematics, assessing bilingual students using flexible linguistic approaches also leads to better outcomes among students who identify as bilingual (Gándara, 2017). Overall, bilingual assessments could improve practice beyond double monolingual approaches, yet more evidence about their potential and appropriateness is needed.

3.Context

3.1. Education in the DRC

The Democratic Republic of the Congo (DRC) is a large country located in Central Africa that has a population of almost 90 million people (The World Bank, 2021). The DRC gained its independence from Belgium in 1960 after 52 years of colonial history. The country is divided into 25 provinces and the city of Kinshasa (capital) and characterized by rich ethnolinguistic diversity (Trapido, 2015). The DRC is a low-income country with its economy depending largely on the mining sector (The World Bank, 2021). Living conditions in the DRC can be harsh, especially for children, many of whom are malnourished (43%) and on average, receive only 4.5 years of learning-adjusted schooling (The World Bank, 2021). Notwithstanding, the DRC has made significant progress in educational outcomes during the past decade. For example, primary completion rate rose to 70% in 2014, and public financing of education has increased from 9% of government expenditure to 18% (Global Partnership for Education, 2021). With the support from donors and non-governmental organizations, the country aims to continue making progress in its quest to provide inclusive and quality education to all.

The education system in the DRC depends on four different ministries: the Ministry of Primary, Secondary, and Initiation to New Citizenship (*Ministère de l'enseignement primaire, secondaire et de l'Initiation à la Nouvelle Citoyenneté*), the Ministry of Higher Education (*Ministère de l'enseignement supérieur et universitaire*), the Ministry of Technical and Vocational Education (*Ministère de l'enseignement technique et professionnel*), and the Ministry of Social Affairs (*Ministère des affaires sociales*), which oversees non-formal education (Ministère de l'Enseignement Primaire Secondaire et Initiation à la Nouvelle Citoyenneté, 2015). The system recognizes three types of schools: '*Ecoles Publiques Non Conventionnées*' which are funded and run by the government, '*Ecoles Publiques Conventionnées*' which are publicly funded but run by religious groups and/or non-governmental organizations, and

'Ecoles Privées' which are privately owned and run (Mokonzi et al., 2020). Most of the schools in the DRC (64%) are 'Ecoles Publiques Conventionnées,' highlighting the large influence that the catholic missionaries have had in the country's education system since the nineteenth century (Mokonzi et al., 2020; Depaepe, Simon, & Vinck, 2015).

In the DRC, free primary education is a constitutional right (Journal Officiel de la RDC, 2011). Yet, this does not automatically translate into all students accessing and completing primary education. Current estimates show that 3.5 million children are out-of-school, most of them in rural areas (Global Partnership for Education, 2021). Primary completion rate is currently 62.4% and secondary enrolment rate is 48% (Ministère de l'Enseignement Primaire Secondaire et Initiation à la Nouvelle Citoyenneté, 2015), indicating that a large portion of students do not finish primary school and many among those who do, decide to drop out after completing that cycle. In addition, learning outcomes in the DRC tend to be very low (see section 3.4). Consequently, a large portion of the DRC's efforts are focused on improving the quality of primary education. Some of the specific actions promoted by the government include: fostering attendance to preschool, ensuring that students start school at the right age, reducing the "indirect" costs for parents to educate their children, focusing on what happens inside the classroom, and building more schools in remote areas to increase the access to schooling for rural students (Ministère de l'Enseignement Primaire Secondaire et Initiation à la Nouvelle Citoyenneté, 2015).

3.2. Language-in-education policies in the DRC

Debates around language-in-education³ in the DRC have existed since colonial times. The discussion goes back to when catholic missionaries were determining the best language to "civilize" and evangelize local Congolese communities (Meeuwis, 1999). The bilingual status of Belgium, the linguistic debates taking place in that country at that time, and the large presence of Flemish missionaries, exerted a large influence on language debates in the colonial DRC. Most missionaries opposed the idea of using French, partly because they believed that French language and culture had corrupted Flanders (Meeuwis, 1999). They also opposed French because it was a suboptimal choice to effectively integrate Catholic values into Congolese society. A more effective approach, according to them, was to use languages that locals understood and used (Depaepe, Simon, & Vinck, 2015). Flemish missionaries advocated for the use of local languages in early grades.

Post-colonial debates around language-in-education are framed in terms of national or *indigenous languages* (i.e., African languages) versus foreign or international languages. After gaining independence, French was installed as the language of instruction for the entire education system in the DRC (Depaepe & Kikumbi, 2018). The rationale was that French would improve the opportunities for the Congolese children, and education in African languages would disenfranchise the population from higher socioeconomic aspirations (Meeuwis, 2011). Later, with the rise of Mobutu and the expansion of the idea of *authenticity*, National Languages were elevated to official languages of instruction until grade 4 (Nthawakuderwa, 1985). French continued to be the official language from grade 5 onwards. The four National Languages recognized as languages of instruction were Lingala, Kiswahili, Ciluba, and

³ Language issues in education

Kikongo (Nthawakuderwa, 1985). These languages are lingua franca, or languages that are used to connect people of different mother tongues, commonly trade languages (Samarin, 1987). Therefore, in many cases, these languages are students' second languages, and not mother tongues or first languages. Among the four, the language that occupies a higher status is Lingala, which is the lingua franca of provinces close to the Equator, and the language of Kinshasa, the DRC's capital (Rassool, 2014). Lingala is also the language of the army and is used in the media, and is therefore, considered the *national lingua franca* (Rassool, 2014).

French remains the official language of the DRC (Journal Officiel de la RDC, 2011). This is at odds with the reality in the country, as it is estimated that only about one third of the population speaks some French and only ten percent is proficient in it (Aber et al., 2016). The constitution of the DRC also recognizes and promotes the four National Languages. But it does not speak to the languages that should be used in education, and the government allows students in grades 1 and 2 to learn in national languages – or local languages, if feasible. This is pedagogically appropriate as most children in the DRC have better understanding of national languages than of French (Gibson, 2019), although, it is logistically challenging as teaching and learning materials in these languages are still unavailable or in an early developmental stage (Louge, Gándara, Gibson, & Thomas, In Press). The current education strategy acknowledges the challenges of teaching in national languages and has increased its emphasis on improving the ability of teachers to teach in languages other than French (Ministère de l'Enseignement Primaire Secondaire et Initiation à la Nouvelle Citoyenneté, 2015).

3.3. Assessments in the DRC

The DRC relies on large-scale assessment data for multiple purposes. There are two well established nationally led large-scale assessments administered to students. The *Test National de Fin d'Études Primaires* (TENAFEP) is a certification exam administered at the end of primary school that students need to pass to advance to secondary education (The World Bank, 2015). In turn, the *Examen d'Etat* marks the end of general track secondary education and is used alongside school grades to determine if a student is eligible to obtain their national certificate of completion of secondary education (*Diplôme d'État*), which is needed to access higher education (The World Bank, 2015). The primary purpose of these exams is certification; however, they are also used to get a sense of students' proficiency on the tested subjects. Currently, the Government is contemplating an additional certification exam half-way through secondary education and modifying the certification system for the entire cycle, including making changes to the *Examen d'Etat* (Ministère de l'Enseignement Primaire Secondaire et Initiation à la Nouvelle Citoyenneté, 2015). Furthermore, in 2021 the DRC launched its first version of the national level monitoring assessment called *Evaluation Nationale CIEAS* (Ministere de l'Enseignement Primaire, Secondaire et Technique, 2021). This sample-based evaluation aims to capture essential skills in literacy and numeracy acquired by second and fourth graders, to support the DRC's quality assurance program.

In recent years, the DRC has also implemented aid-supported large-scale assessments to monitor early grade learning outcomes at the national level. In particular, the country has used the Early Grades Reading Assessment

(EGRA) and the Early Grades Mathematics Assessment (EGMA) to inform the current education sector strategy. For example, the poor results on the EGRA and EGMA studies conducted in 2012 as part of the Project d'Amélioration de la Qualité de l'Éducation (PAQUED) pushed the government to reinforce its focus on learning outcomes and its efforts on closing gender gaps (Ministère de l'Enseignement Primaire Secondaire et Initiation à la Nouvelle Citoyenneté, 2015). More recently, an EGMA study conducted in 2018 as part of the ACCELERE! 1 Project highlighted specific areas that the government should prioritize in the upcoming years and provided specific recommendations for improving learning outcomes, such as adding time to the vocabulary building activities and word explanations in the teaching guides (School-to-School International, 2019). Overall, the government uses these assessments to periodically monitor the status of educational outcomes and gain insights on how to further improve the system.

The DRC also participates in comparative studies, such as the Multiple Indicator Cluster Surveys (MICS) study from UNICEF. The MICS is a large international survey administered to children and women around the world. It includes a module on Foundational Learning (FL) Skills aimed at capturing foundational literacy and numeracy skills for children in and out-of-school (UNICEF, 2017). The country also participates in regional large-scale assessments. In addition, as a member of the Conference des Ministres de l'Éducation des États et Gouvernements de la Francophonie (Confemen), the DRC participates in the Programme d'Analyse des Systèmes Éducatifs de la Confemen (PASEC). The PASEC is an assessment that evaluates national systems of education and produces comparable data in literacy (typically French) and mathematics, among francophone countries (Confemen, 2021). PASEC receives methodological support from many international organizations. Administered every five years, it provides insight onto what works in each educational system and has consistently been used to inform national policy in the DRC. PASEC also provides an opportunity to collaborate with other francophone countries in the region, in their shared quest to improve the educational outcomes for students.

3.4. Literacy outcomes in the DRC

One of the most important global education indicators is Indicator 4.1.1 from the SDGs. The indicator requires countries to estimate the:

“Proportion of children and young people: (a) in grades 2/3; (b) at the end of primary; and (c) at the end of lower secondary achieving at least a minimum proficiency level in (i) reading and (ii) mathematics, by sex” (UIS-UNESCO, 2017).

It is unclear whether the DRC is planning to use the newly established Evaluation Nationale CIEAS 2021 to feed SDG indicator 4.1.1. In the meantime, PASEC data can be used to estimate indicator 4.1.1, as the assessment is aligned with the Global Proficiency Framework. In addition, MICS and EGRA data measure skills that are of relevance to indicator 4.1.1 and can be used to provide alternative estimates to indicator 4.1.1. As it will be shown in the following sections, these estimates are not consistent across assessments, highlighting the inherent dependence of indicator 4.1.1. to methodological choices.

3.4.1. MICS

The MICS FL module targets students by age (7-14 years) and not by grade level. However, MICS FL tasks are targeted at Grade 2 level of difficulty (UNICEF, 2017). The MICS literacy portion comprises two tasks: a) reading a passage with fluency (untimed), and b) an associated set of reading comprehension questions (three literal questions and two inferential questions). Based on the MICS methodology, a child meets minimum competency levels if they a) read correctly 90% of the words in the passage, and b) correctly answer five reading comprehension questions (UNICEF, 2017).

The latest MICS results (2018) show disparities in learning outcomes across urbanicity, province, and sex categories. In the DRC, the MICS FL was conducted in French (INS, 2019). As shown in Table 1, at the country level, there is overwhelming disparity between urban and rural students, with urban students achieving four times the foundational literacy of rural peers. In addition, there are significant differences by province. Students in provinces with larger urban centers (Equateur, Nord Kivu, Haut Katanga, and Kasai Oriental) tend to perform better, and on average, Lingala-phone provinces score lower than the other groups. Disparities between girls and boys vary by urbanicity and province.

Table 1 Percentage demonstrating foundational knowledge in reading, MICS, Children Aged 7-14

| Results by Urbanicity | | | |
|--|-------------|--------------|--------------|
| | Boys | Girls | Total |
| Urban | 15.20% | 15.60% | 15.40% |
| Rural | 4.40% | 3.10% | 3.80% |
| Results by Province⁴ | | | |
| Lingala Provinces | | | |
| Equateur | 7.40% | 1.90% | 4.50% |
| Sud Ubangi | 6.00% | 0.00% | 3.30% |
| Kiswahili Provinces | | | |
| Nord Kivu | 13.00% | 18.90% | 15.50% |
| Sud Kivu | 8.70% | 9.30% | 9.00% |
| Haut Katanga | 9.10% | 10.20% | 10.30% |
| Lualaba | 6.20% | 15.20% | 5.10% |
| Ciluba Provinces | | | |
| Kasai Oriental | 8.30% | 6.40% | 7.50% |
| Kasai Central | 6.10% | 1.20% | 3.70% |

Source: INS, Enquête par grappes à indicateurs multiples, 2017-2018, rapport de résultats de l'enquête. Kinshasa, République Démocratique du Congo.

Table 2 shows the results for students enrolled in grades 2 and 3. The proportion of students achieving foundational literacy is considerably lower for these grade levels than for the overall sample, which includes children as old as 14-year-olds.

⁴ Provinces relevant to this study

Table 2 Percentage demonstrating foundational knowledge in reading, MICS, Grade 2 and Grade 3 students

| | Boys | Girls | Total |
|---------|-------|-------|-------|
| Grade 2 | 1.70% | 1.40% | 1.50% |
| Grade 3 | 5.70% | 3.80% | 4.80% |

Source : INS, Enquête par grappes à indicateurs multiples, 2017-2018, rapport de résultats de l'enquête. Kinshasa, République Démocratique du Congo.

3.4.2. EGRA

The EGRA is an assessment used for tracking foundational literacy in the context of donor-based projects (United States Agency for International Development, 2019). The EGRA measures core skills needed for the acquisition of reading using several tasks, most notably, a timed Oral Reading Fluency (ORF) task and a Reading Comprehension task (Dubeck, Gove, & Alexander, 2016; RTI International, 2015). Typically, the number of reading comprehension questions that a student is asked depends on how far the student reads on the passage used to measure ORF, and ranges between 0 and 5. The EGRA can be adapted into the local context and administered in local languages, yet it is rarely adapted to the linguistic characteristics of local languages (Trudell, 2016). While the EGRA measures foundational skills that are not curriculum specific, EGRA studies target difficulty at the corresponding grade level.

The EGRA was not designed for global monitoring. Although, its extensive use has led to an overreliance on the assessment for defining country-level benchmarks and monitoring country-level literacy outcomes. The EGRA is often used within aid-funded projects to evaluate learning outcomes and since 2010, has been used in large-scale nationally representative evaluations in the DRC. As part of a joint effort between the DRC government, USAID, and program implementers, the EGRA has also been used to establish country-level benchmarks for several skills, including ORF and Reading Comprehension. The benchmarks for ORF in national languages were set between 32 and 35 correct words per minute (cwpm), depending on the language, and the benchmarks for Reading Comprehension were set at answering at least 60% of questions correctly (School-to-School International, 2019). The latest national-level EGRA study corresponds to the one administered in 2018 (see Figure 1). The results of that study show that the proportion of grade 2 students achieving country-level benchmarks for ORF and Reading Comprehension in national languages is remarkably low (see Table 3). While the benchmarks may be too high to begin with, Table 3 shows that the average ORF scores are at best, 9.96 cwpm (correct words per minute), and may be as low as 1.94 cwpm, depending on the provincial group.

Figure 1 National Level EGRA Assessments Administered in the DRC, by Grade Level

| Project | 2010 | 2011 | 2012 | 2013 | 2014 | 2015 | 2016 | 2017 | 2018 |
|------------|--------|------|------|------|--------|------|------|------|------|
| PAQUED | 2,4,6* | | | | 2,4,6* | | | | |
| ACCELERE!1 | | | | | | 2* | | 2* | 2* |

* And equivalent grades for students in Accelerated Learning Programs (ALP or CRS).

Table 3 EGRA 2018 Results – by Provincial Group

| | Average ORF – cwpm | | Percentage of Students Meeting Benchmarks (Public Schools) | |
|-----------------|--------------------|------|--|-----------------------|
| | Public Schools | ALPs | ORF | Reading Comprehension |
| Kiswahili-phone | 5.03 | 9.96 | 5.36% | 1.21% |
| Lingala-phone | 1.94 | 6.4 | 1.66% | 1.34% |
| Ciluba-phone | 3.67 | 6.23 | 2.93% | 1.79% |

Source: STS 2019

3.4.3.ACCELERE! Quarterly Assessments – Grade 2

The ACCELERE! Quarterly Assessments (QAs) were curriculum-based assessments administered within the context of the ACCELERE! 1 program to grade 1, 2 and 3 students. The QAs were fundamentally different to an EGRA because they were closely aligned to the ACCELERE! Program (as opposed to measuring general literacy skills). As such, the QAs included tasks that were different to those found in the EGRA toolkit and were tailored to the program and population assessed. However, the QAs also included EGRA-like tasks, namely, a letter-identification task, an oral reading fluency task, and a reading comprehension task. In particular, the QAs administered in 2017-2018 captured Oral Reading⁵ and Reading Proficiency among grade 2 students and grade 2 equivalent students attending non-formal schools (ALPs), at the end of the academic year. The ORF passages targeted grade 2 students and were adapted from textbooks used in the program, administered in national languages and untimed. The reading comprehension questions included a mix of literal and inferential questions and were administered to all students. Differently to the EGRA, students were allowed to look at the passage as they answered the reading comprehension questions. The sample of students used was similar to the sample used in the ACCELERE! EGRA 2018 study.

The results on the 2017-2018 Quarterly Assessments (see Table 4) show higher mean figures than those in the EGRA 2018 study, with students reading between 9.03 correct words in Sud Ubangi (out of 42) and 23.48 correct words in Nord Kivu (out of 41). Because the number of words varied by language group, the results for Oral Reading in Table 4 are also presented as percent correct. Based on this measure, students were able to correctly read between 21.5% and 57.3% of the words in their corresponding passages.

Table 4 Quarterly Assessment Results – Oral Reading, Grade 2 - 2018

| | Oral Reading Fluency (Average Correct Words) | | | Oral Reading Fluency (Percent Correct) | | |
|--------------|--|-------|-------|--|-------|-------|
| | Girls | Boys | All | Girls | Boys | All |
| Equateur | 15.51 | 17.25 | 16.39 | 36.9% | 41.1% | 39.0% |
| Sud Ubangi | 8.96 | 9.10 | 9.03 | 21.3% | 21.7% | 21.5% |
| Nord Kivu | 22.65 | 24.47 | 23.48 | 55.2% | 59.7% | 57.3% |
| Sud Kivu | 15.51 | 17.46 | 16.53 | 37.8% | 42.6% | 40.3% |
| Haut Katanga | 17.18 | 20.71 | 18.96 | 41.9% | 50.5% | 46.2% |

⁵ The Oral Reading Fluency passage used at that point was the same as the one used in this study

| | | | | | | |
|----------------|-------|-------|-------|-------|-------|-------|
| Lualaba | 16.07 | 20.16 | 18.08 | 39.2% | 49.2% | 44.1% |
| Kasai Oriental | 13.74 | 15.71 | 14.73 | 40.4% | 46.2% | 43.3% |
| Kasai Central | 10.25 | 11.15 | 10.69 | 30.2% | 32.8% | 31.4% |

The results for reading comprehension (see Table 5) show that between 7.5% (Sud Kivu) and 32.2% of students (Kasai Oriental) responded to at least 60% of the questions correctly.

Table 5 Quarterly Assessment Results – Reading Comprehension, Grade 2 - 2018

| | Reading Comprehension (Correct out of 4) | | | Percentage of Students Meeting Reading Comprehension Benchmarks (≥ 3 correct questions) | | |
|----------------|--|------|------|--|-------|-------|
| | Girls | Boys | All | Girls | Boys | All |
| Equateur | 1.39 | 1.62 | 1.50 | 25.6% | 31.9% | 28.8% |
| Sud Ubangi | 0.95 | 0.92 | 0.93 | 13.6% | 14.5% | 14.0% |
| Nord Kivu | 1.33 | 1.30 | 1.32 | 26.2% | 21.7% | 24.2% |
| Sud Kivu | 0.61 | 0.54 | 0.57 | 8.1% | 7.0% | 7.5% |
| Haut Katanga | 1.02 | 1.21 | 1.12 | 15.8% | 23.5% | 19.7% |
| Lualaba | 0.86 | 1.07 | 0.96 | 12.8% | 17.5% | 15.1% |
| Kasai Oriental | 1.37 | 1.66 | 1.51 | 28.0% | 36.3% | 32.2% |
| Kasai Central | 1.14 | 1.26 | 1.20 | 23.6% | 26.6% | 25.1% |

3.4.4. PASEC

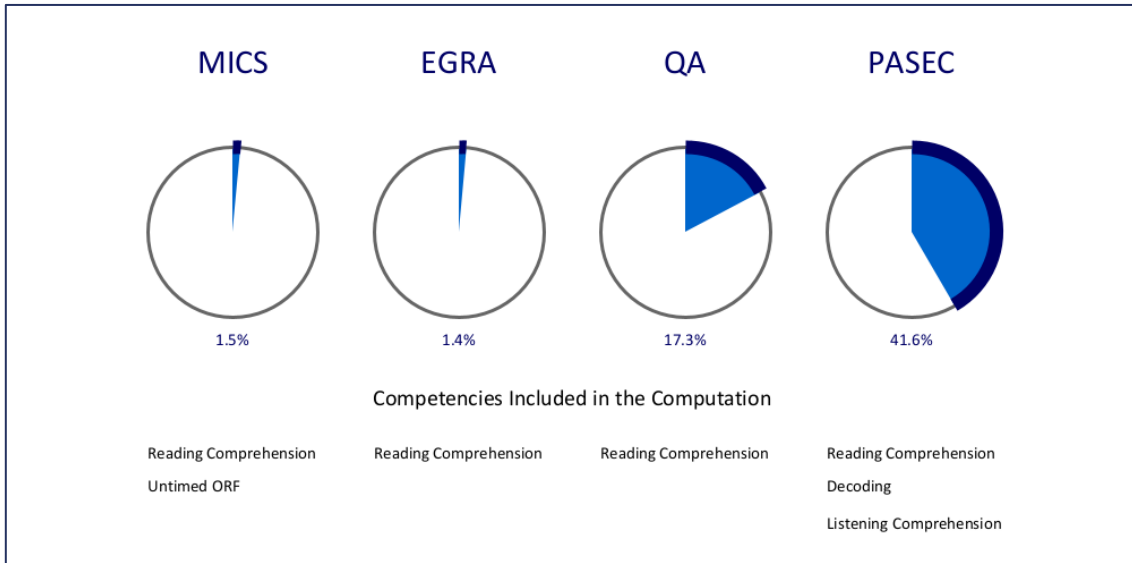
The PASEC tests are administered at the end of grade 2 and grade 6, at the end of primary school. The literacy portion of the grade 2 tests measures three domains – listening comprehension, decoding (letters, syllables, and words), and reading comprehension. The literacy portion of the end of primary test measures comprehension of words and phrases, and reading comprehension, with the latter accounting for most of the scores (Confemen, 2020). The PASEC scores are scaled (e.g., 0 – 610 points) and classify students into several levels of proficiency sets after a standard setting procedure. The PASEC provides a benchmark of *sufficient proficiency*, based on psychometric and curricular analyses (Confemen, 2020).

Based on the last PASEC (2019), the DRC ranks among the lowest performing countries, with 41.6% of grade 2 students and 27.1% of grade 6 students, at the end of primary schooling, meeting standards of sufficient proficiency.

3.4.5. Proficiency across assessments

The data provided in sections 3.4.1 - 3.4.4 constitutes just a sample of data that can be used to gain insight into the proficiency of students in the DRC. As noted, the proficiency outlook varies with regards to the assessment used. The disparities in proficiency levels and outcomes highlight the dependence between performance and instrument, as well as the criteria for proficiency. In the absence of linked assessments and common proficiency levels, data from different assessments are simply not comparable, even when measuring overlapping skills (see Figure 2).

Figure 2 Percentage of Students Meeting Expectation using Different Assessments and Proficiency Criteria, Grade 2 Students



To use these assessments toward informing indicator 4.1.1, scores would need to be *linked*. Linking would place the scores of two assessments that may differ in content and difficulty on a same scale. However, the resulting scores are not necessarily comparable in terms of interpretation, and such determination lies in the scope, content, and specifications of the assessments. For assessments capturing similar constructs, linking provides a transformation function so that scores from one test or assessment can be translated into another. Linking assessments such as the EGRA and PASEC could allow practitioners to estimate indicators such as 4.1.1 using data that do not come from the same instrument. However, high-stakes uses of test scores such as global monitoring, would demand a high degree of similarity between the assessment frameworks and a very precise linking procedure⁶.

Several properties of the assessments and scores contribute to precise and meaningful linking. For example, all the assessments reviewed in this section capture some measure of reading comprehension, but such tasks tend to consist of a few questions only; for example, the EGRA typically uses a 5-item reading comprehension task. A linking procedure established on a task that has a maximum of 6 score options⁷ is likely to be unreliable. A more reliable linking procedure should be based on more data points. For example, the PASEC and the EGRA test overlap in a few tasks: reading comprehension, word reading, letter identification, and listening comprehension. Using all those tasks could provide a more reliable basis for a linking procedure. However, EGRA tasks such as word identification and letter identification are timed, and therefore, not directly comparable to untimed versions of the tasks. In addition, the EGRA reading comprehension task is such that the number of questions attempted is contingent to the performance on the ORF task, limiting its comparability to other forms of reading comprehension tasks. In sum, the possibility of linking assessments, even when they measure similar tasks, is contingent upon other characteristics such as length and administration procedures.

⁶ For a full discussion on linking, check Kolen and Brennan (2014).

⁷ Using Classical Test Theory

Ideally, linking the scores from the EGRA and PASEC would entail aligning the content (e.g., ensuring that the listening comprehension questions capture the same type of information), administration of tasks (e.g., timed, or untimed), and the scoring methods (e.g., total scores based on all the tasks). Alternatively, as proposed by the UNESCO Institute of Statistics (UIS) and the United States Agency for International Development (USAID), the assessments could be linked by means of a non-statistical method known as *policy linking* (Management Systems International, 2019). Policy linking consists of aligning categories of performance using comparable definitions of proficiency, provided by a common framework such as the Global Proficiency Framework (GPF) (USAID, 2019). In particular, using EGRA data and the GPF, practitioners and policy makers could establish new benchmarks to classify students into categories of proficiency that are aligned with those reported by the PASEC, whose framework mirrors the global framework. More research is needed to understand the extent to which this methodology is appropriate in a high-stakes context.

4.Data and methodology

4.1.Data

Data for this study comes from a large-scale literacy program in the DRC that started in 2015 and was scheduled to end in 2021⁸. The dataset contains 2,395 records of grade 3 students enrolled in public schools or grade 3 equivalent students enrolled in Accelerated Learning Programs (ALPs), collected at the end of the second trimester of the 2018/2019 academic year (March 2019). The records contain demographic information on the students and teachers, school data such as location, classroom information such as size, and the scores of the students on short literacy assessments called Quarterly Assessments (QAs). A sex-balanced sample of students was selected using a two-stage stratified clustered methodology.

Table 6 Characteristics of the Dataset

| Province | Students | % ALP | % Girls |
|----------------|--------------|------------|------------|
| Equateur | 320 | 37% | 47% |
| Haut-Katanga | 342 | 37% | 49% |
| Kasai Central | 253 | 49% | 44% |
| Kasai Oriental | 345 | 39% | 50% |
| Lualaba | 350 | 27% | 49% |
| Nord Kivu | 251 | 100% | 52% |
| Sud Kivu | 207 | 100% | 48% |
| Sud Ubangi | 327 | 34% | 50% |
| Total | 2,395 | 49% | 49% |

⁸ Ministry approval of the project included a revision of ethical considerations for data collection and use. All students and teachers in the sample provided consent.

The ACCELERE! QAs measured essential literacy skills taught during each trimester of instruction. The ACCELERE! curriculum was structured around essential skills taught each trimester in National Languages and/or French, depending on the grade level. From grade 3 onwards, students learned both French and National Languages. The third-grade curriculum built upon the interdependence hypothesis described earlier, with some literacy skills acquired in National Languages (grades 1 and 2) being used to develop skills in French. The third-grade curriculum also assumed that some literacy skills would be learned simultaneously across languages. The Quarterly Assessments for grade 3 focused on skills in National Languages (broadly assumed to be L1s) and French (broadly assumed to be an L2), that are not necessarily the same skills. The development of the assessments followed an evidence-based iterative process. The essential literacy skills were identified by experts who had in-depth knowledge of the teaching and learning materials. Tasks meant to capture these essential skills were drafted and piloted, and only those with good psychometric qualities were retained. The QAs were different by trimester, yet they all included Bilingual Tasks that captured bilingual and/or metalinguistic skills, depending on the trimester. Doing so allowed the team to examine the transference of skills between tested languages and their co-development.

The assessment used in this study had four tasks, focusing on the acquisition of high-frequency words (French), oral reading fluency (National Languages), decoding (French), and writing (Bilingual). Measuring the acquisition of high frequency words and decoding is consistent with methodologies to teach opaque languages such as French; measuring oral reading in National Languages is consistent with research that shows its relevance to reading comprehension; and measuring written production across languages is consistent with the developmental expectations for grade 3 students and with research that shows that bilingual students could transfer metalinguistic skills (such as knowing that sentences begin with capital letters, etc.) while developing their writing skills in more than one language. Due to the uneven development of National Languages and French, as well as their linguistic differences, it did not make sense to measure the same skills across languages. More details about the assessment can be found in Table 7.

Table 7 Tasks in the ACCELERE! Quarterly Assessments

| Task | Description | Number of Items |
|---------------------------|--|------------------------|
| High Frequency Words | Students had to read out loud common French words | 20 |
| Oral Reading Fluency | Students had to read out loud a grade-level passage | 34 - 42 |
| Decoding – Parts I and II | Students were given cards with printed French words and had to create groups with words that began with the same sounds (Part I), or ended with the same sound (Part II) | 7 words |

| | | |
|----------------------------------|--|----------------------------------|
| Decoding – Parts III and IV | Students listened to three words aloud and had to identify the pair that began with the same sound (Part III), and that ended with the same sound (Part IV). | 4 items per part |
| Sentence Analysis and Production | Students had to write down phrases in the corresponding National Language (Part I) and in French (Part II). Analytical scores were provided for both, the ability to write full words and to use written conventions across languages (e.g., capital letter, spaces between words and final points). | 2 sentences in National Language |

4.1. Analytical Framework

4.1.1. First research question – Challenges faced by bilingual education

To respond to the first research question, a literature review on bilingual education within the Sub-Saharan African region was conducted. The review was centered around the challenges faced by bilingual education programs, with a focus on primary and secondary school. The review was exploratory (Gheondea-Eladi, 2015) and provided an overall summary of challenges to bilingual education in Sub-Saharan Africa. The identification of literature followed a structured and replicable approach. The review focused on peer-reviewed articles written from 2011 onwards. The first step consisted of conducting three different searches in Google Scholar and identifying the 30 most relevant articles (as per Google Scholar metrics). The first search included the following keywords: a) mother tongue, b) education, c) multilingualism, d) policy, e) Sub-Saharan Africa, and f) DRC, in addition to relevant synonyms for these keywords⁹. The second search included the following keywords: a) mother tongue, b) education, c) multilingualism, d) policy, e) Sub-Saharan Africa, f) DRC, g) socio-political challenges, h) system challenges, and i) logistical challenges, in addition to relevant synonyms for these keywords¹⁰. The third search included the following keywords: a) mother tongue, b) education, c) multilingualism, d) policy, e) Sub-Saharan Africa, f) DRC, and g) intersectionality, in addition to relevant synonyms for these keywords¹¹. To increase the number of articles, the search was replicated by removing certain keywords and/or synonyms.

The second step in the literature review consisted of retaining those articles that were relevant to the topic of *challenges of bilingual education in Sub-Saharan Africa and/or the DRC*. From the initial search, documents were removed if: a) they were not peer-reviewed articles; b) they were written in a language other than English, Spanish or French¹²; c) they did not pertain to education; d) they were situated in contexts that excluded Sub-Saharan Africa; e)

⁹ Full set of keywords: mother tongue, L1, first language, second language, education, school, primary, secondary, multilingualism, bilingualism, policy, law, Sub-Saharan Africa, DRC, Democratic Republic of the Congo

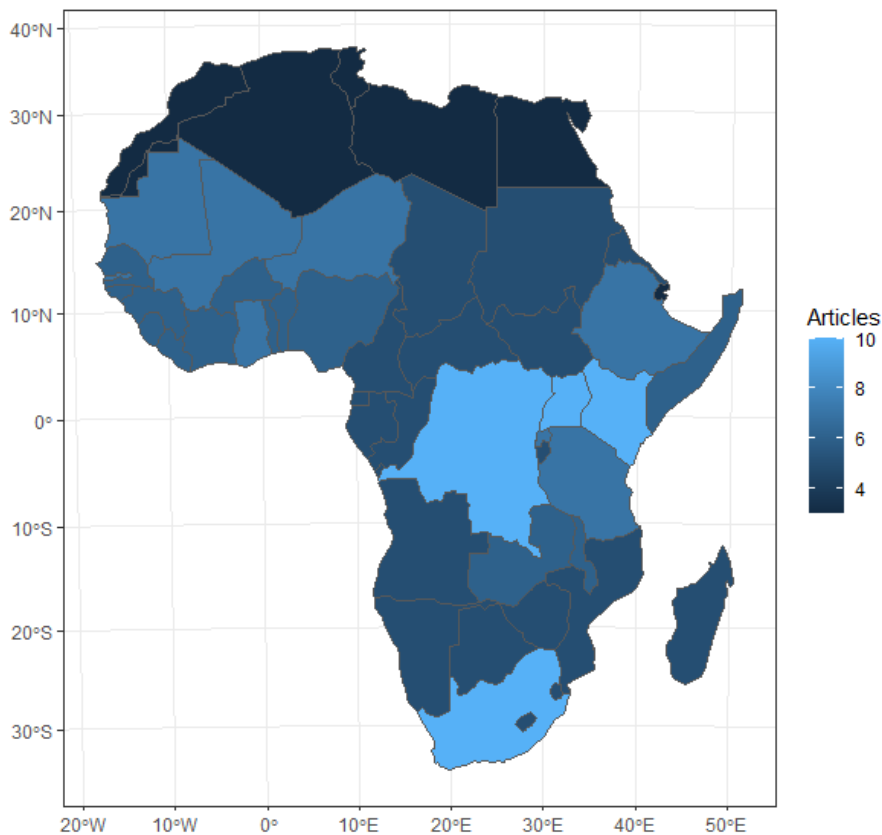
¹⁰ Full set of keywords: mother tongue, L1, first language, second language, education, school, primary, secondary, multilingualism, bilingualism, policy, law, Sub-Saharan Africa, DRC, Democratic Republic of the Congo, socio-political challenges, challenges, logistical challenges, resource challenges, system challenges

¹¹ Full set of keywords: mother tongue, L1, first language, second language, education, school, primary, secondary, multilingualism, bilingualism, policy, law, Sub-Saharan Africa, DRC, Democratic Republic of the Congo, intersectionality, migration, ethnicity, race, and gender

¹² Languages accessible to the author

they focused on higher or adult education; f) their full-texts were unavailable to the author; g) they did not add new information; or h) they were simply not relevant for the review. The final number of articles was narrowed down to 34 and covered different areas within Africa, particularly, Sub-Saharan Africa (see Figure 3).

Figure 3 Final Selection of Articles by Geographic Focus



4.1.2. Second research question – Representation of bilingual literacy

To respond to the second research question, this study used Structural Equation Modelling (SEM) to evaluate the latent structure that best represented bilingual data collected via the Quarterly Assessment. Four models were fitted on the dataset. First, a model which included a single factor accounting for the responses to French, National Language, and Bilingual Tasks was used. This model assumed a unidimensional latent structure underlying the data. The rationale to include this model was to test the extent to which French and National Languages were undistinguishable in terms of construct representation. This model is theoretically aligned to the unified view on languages (see section 2.3). Second, a bidimensional model was applied, which comprised a Bilingual Literacy factor measured directly by bilingual and metalinguistic parcels and impacting two other latent variables – French Literacy and National Languages Literacy. The second model recognizes the influence of Bilingual Literacy on each language and was included in accordance with some of the assumptions underlying the ACCELERE! Curriculum. To be clear, the second model assumes that Bilingual Literacy impacts literacy in French and National Languages, rather than confirming a causal relationship between the two. Third, a tridimensional model was fitted, which included a second

order factor – Literacy – and three orthogonal first order factors – French Literacy, National Languages Literacy, and Bilingual Literacy. This model also allowed the bilingual and metalinguistic parcels to load directly onto the Bilingual Literacy factor. The rationale to include this model was to evaluate whether students’ literacy was inherently multidimensional, without making assumptions about the relationship between Bilingual, French, and National Languages Literacy. Fourth, a model with two first order orthogonal factors – French Literacy and National Languages Literacy – was applied. These two factors accounted for the variance of all items, including the bilingual and metalinguistic parcels¹³, and were forced to be uncorrelated. The rationale to include the fourth model was to test whether literacy in French and National Languages were essentially independent. All models used parcels as opposed to individual items. Parcels were used to reduce the very large number of items (more than 50 per form) to a more manageable set. Parcelling can be effective if the grouped items are reasonably unidimensional (Marsh et al., 2013). Therefore, prior to parcelling, a series of exploratory factor analyses were conducted to ensure that the sets of items were unidimensional and that grouping them did not result in unintended consequences such as masking sources of misfit.

Table 8 Models and Interpretation

| Model | Interpretation if Best Fitting Model |
|----------------------------|--|
| Model I - Unidimensional | Literacy among bilinguals is best represented as a unidimensional construct. There is no need to distinguish between literacy in French and in National Languages. |
| Model II - Bidimensional | Literacy in French and National Language are best represented as a bidimensional construct being influenced by an overarching Bilingual Literacy factor. |
| Model III – Tridimensional | Literacy among bilinguals is best represented as a tridimensional construct, where French, National Language, and Bilingual parcels are measured separately. |
| Model IV – Orthogonal | Literacy in French and National Language are uncorrelated and best represented by two independent latent variables |

Evaluating the appropriateness of a SEM model requires integrating theory and statistical parameters, such as fit indices and factor loadings obtained in each solution (Smith & McMillan, 2001). A model should be grounded in theory, yield relations between variable that are expected or plausible, and fit the data reasonably well. A wealth of indices exists to evaluate SEM solution. Among the mostly used indices are the Chi-square (χ^2) test, the Root Mean Square Error of Approximation (RMSEA), the Comparative Fit Index (CFI), and the Tucker-Lewis index (TLI), each associated with unique strengths and limitations. For example, Chi-square tests are too sensitive to sample size and a large sample size may lead to a rejection of an adequate model (Smith & McMillan, 2001). Or RMSEA, CFI, and TLI

¹³ These parcels were forced onto the National Languages’ factor, as students were more proficient in these languages than in French

may be sensitive to different types of model misspecification (Shi, Lee, & Maydeu-Olivares, 2019). Therefore, fitted indices were not interpreted individually but evaluated as a whole. Models also need to be analysed in terms of local solutions such as factor loadings and covariances. For example, the strength and direction of factor loadings is relevant in understanding the quality of a solution. In this regard, factor loadings below 0.4 were considered low, and factor loadings above 0.8 were considered high (Shi, Lee, & Maydeu-Olivares, 2019). All these considerations were used to identify the best model.

Table 9 Fit Indices and Interpretation

| Type of Index | Index | Description | Threshold Values |
|----------------------|---|---|--|
| Absolute Fit | Chi-square (χ^2) | Classic index of fit which compares the covariance input matrix with the estimated one. | A non-significant χ^2 test is indicative of good model fit (Smith & McMillan, 2001) |
| Parsimony Correction | Root Mean Square Error of Approximation (RMSEA) | Average discrepancy between the correlations in the input matrix and those predicted by the solution, corrected by parsimony. Ranges between 0 (perfect fit) and 1. | 0= perfect fit, < 0.05 = close fit, 0.05 to 0.08 = fair fit, 0.08 to 0.1 = mediocre fit, and > 0.1 = poor fit (Smith & McMillan, 2001) |
| Comparative Fit | Comparative Fit Index (CFI) | Based on the difference in absolute fit between the solutions. It ranges between 0 and 1 (good fit). | Acceptable solution is > 0.9 (Smith & McMillan, 2001) or above > 0.95 (Brown, 2015) |
| | Tucker-Lewis index (TLI) | Like CFI but compensates for model complexity. It ranges between 0 and 1 (good fit). | Acceptable solution is > 0.95 (Brown, 2015) |

4.1.3. Third research question – Equivalence of bilingual assessments

To respond to the third research question, the study examined whether the best fitting model was *invariant* across students of different linguistic backgrounds. Invariance refers to the equivalence of the scores across subgroups. Invariance between subgroups can be interpreted as the lack of *bias*. Bias refers to the nuisance factors that prevent an appropriate comparison between different subgroups (van de Vijver & Leung, 1997). Three levels of measurement invariance were examined: *structural*, i.e., whether the constructs are equivalent across subgroups, even though not operationalized in the same way; *metric*, i.e., when in addition to structural equivalence, the units of measurement are the same across groups; and *scalar*, i.e., when in addition to metric equivalence the origins of the scale are the same (and therefore, scores can be fully compared) (van de Vijver & Leung, 1997). When scores are metric equivalent, differences between scores can be compared; when scores are scalar equivalent, means across subgroups can be compared.

The datasets used in this study contained self-reported information on how often students spoke French, National Languages, and/or Minority Languages at home. Based on this information and on the distribution of students, two subgroups were created. First, students were classified with regards to the linguistic diversity of their home environments and the number of languages they spoke at home. In particular, two groups were created – students who spoke one language at home vs students who spoke at least two languages at home. Second, students were classified with regards to whether they spoke a minority language¹⁴ at home. Using multigroup models, the study evaluated the equivalence of the models across these two subgroups of students using different levels of constraints.

Figure 4 Proportion of Students by Number of Languages Spoken at Home

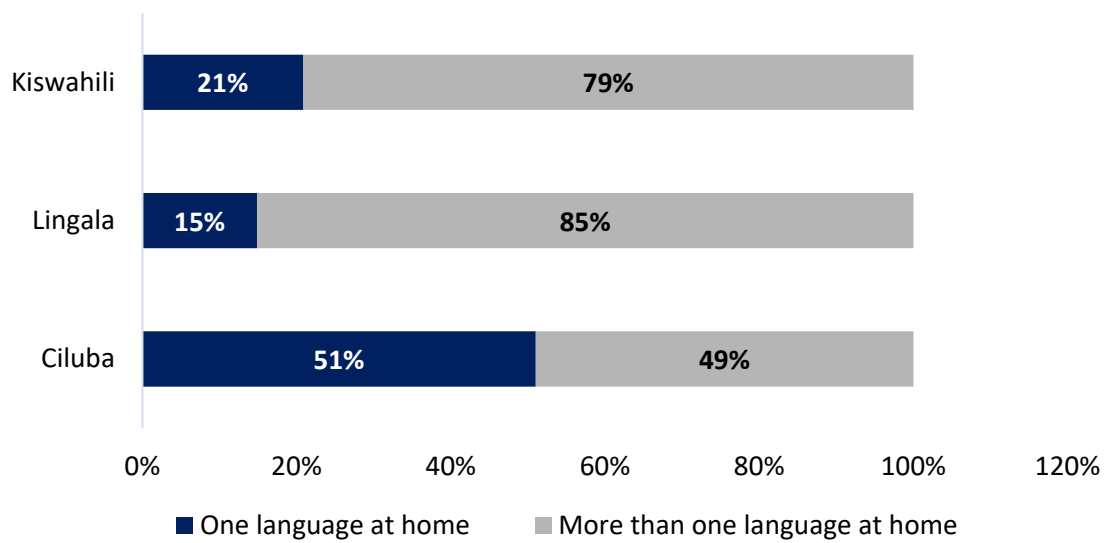
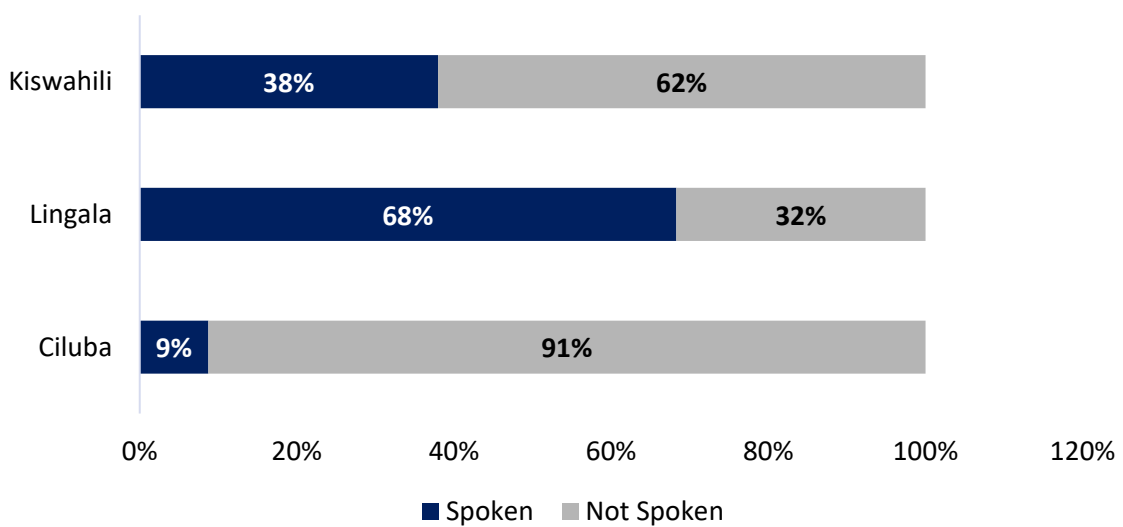


Figure 5 Proportion of Students by Minority Language Spoken at Home or Not

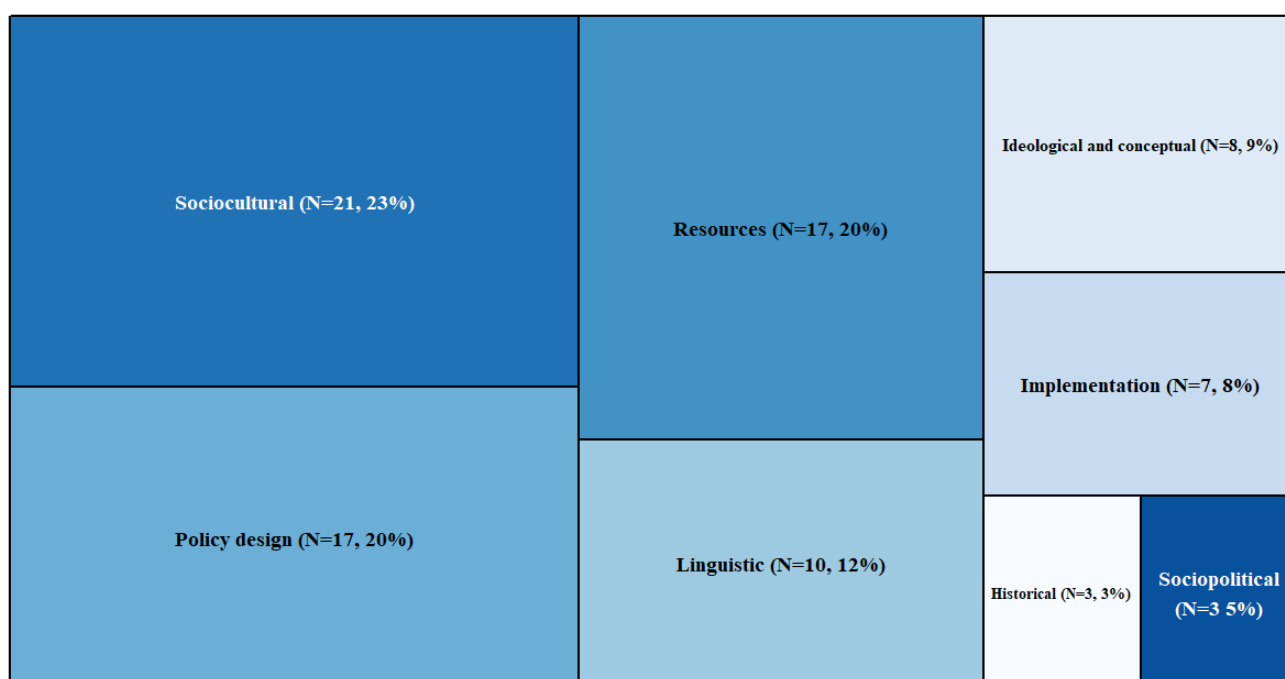


¹⁴ Other than French, Ciluba, Lingala or Kiswahili

5. Literature review findings

A systematic review of the literature indicated that bilingual education programs in Sub-Saharan Africa face primarily seven types of challenges: a) policy design challenges, b) resource challenges, c) sociocultural challenges, d) ideological and conceptual challenges, e) linguistic challenges, f) historical challenges, and g) implementation challenges. A classification of the main challenges discussed in each document¹⁵ showed that the most predominant of all are sociocultural, resources, and policy design challenges, which are the primary focus of at least twenty percent of the articles (see Figure 3).

Figure 6. Predominance of Challenges in the Reviewed Documents



5.1.1. Sociocultural challenges

Sociocultural challenges referred to those ingrained in the social and/or cultural characteristics of the contexts in which bilingual education operates. In this review, sociocultural challenges included ethnolinguistic diversity, identity valorisation, and intersectionality.

Ethnolinguistic diversity

Multiple authors identify the ethnolinguistic diversity of the region as an inherent challenge to the success of bilingual education. Sub-Saharan Africa, and Africa more generally, are marked by a high yet unevenly distributed ethnolinguistic diversity, with many spoken languages (at least forty per country) and by diasporas (Ekkehard, 2018; Rassool, 2014). The languages spoken in any given place may change over time and are difficult to identify.

¹⁵ More than one challenge could be discussed per document

Consequently, making choices such as which language(s) to use in schools can become very challenging (Babaci-Wilhite, 2013; Igboanusi, 2014). The idea of teaching every child in their mother tongue, in a pedagogically optimal way, can be unrealistic (Nyarko Ansah, 2014). As a result, ethnolinguistic diversity poses challenges to bilingual education, and the predominance of foreign languages is sometimes blamed for or justified on the degree of multilingualism of the African nations (Bokamba, 2011).

Identity valorisation

The literature review highlighted the challenges from language-in-education policies affecting people's identities and their sense of belonging. Languages have currency, and linguistic choices are not neutral but rather contingent upon the status of languages within a given society (Babra Gora, 2017; Rassool, 2014). Language choices made by individuals are based on the contexts in which they function and their perceptions around languages (Muaka, 2011). Similarly, choices made in the educational context respond to and affect those perceptions, such as the symbolic power of a language. Language-in-education policies may be used, for example, to legitimize the language(s) of the dominant groups (Nishioka & Durrani, 2018; Orwenjo, 2012) and those that do not recognize the languages that people use in their daily lives can have detrimental effects on identity conflicts (Muaka, 2011). In particular, the choice of languages used in education will matter to learners who face the question of whether their identities are valued and validated (Babaci-Wilhite, 2013). The languages used in education also matter to the identity of teachers and other school personnel (Sibomana, 2020), especially in contexts where languages are quasi ethnic identifiers (Borg, 2015). Africa's multilingualism is characterized by polyglossia, where foreign languages hold a higher status than indigenous languages (Ekkehard, 2018). African or indigenous languages may even be associated with backwardness and underdevelopment, despite being used in the everyday context (Ekkehard, 2018). Therefore, current models of bilingual education that emphasize foreign over indigenous languages, are immensely complex from the identity point of view. Language policies in education can act as power signals and contribute to the establishment and/or maintenance of language hierarchies and their corresponding social inequalities.

Intersectionality

The review also highlighted the complexity of sociocultural challenges, which tend to be intersectional in nature. The nature and the impact of language policies on learners' identities are moderated by other characteristics, for example, if they live in rural areas or if they are immigrants or refugees. Exploring the urban-rural divide is important for understanding and framing of language-in-education policy in Sub-Saharan Africa. Rural students tend to live in more isolated areas and typically belong to lower socioeconomic status, but generally show more group cohesion (Evans, 2018). Often, rural areas reflect what Ekkehard (2018) defines as multi-monolingualism, in which multilingualism can be seen as a joint collection of monolingual pockets. As such, rural and urban contexts present different challenges to bilingual education and contribute to different policy sets. For example, in Kenya, the language of instruction in the first years may be English, Kiswahili, or an indigenous language, depending on the

urbanicity and linguistic cohesion of the area in which a school is located (Muaka, 2011). Specifically, English is the language of instruction in urban areas, Kiswahili is the language of instruction until grade 3 in multi-ethnic rural areas, and indigenous languages are used until grade 3 in linguistically homogenous rural areas (Muaka, 2011). Another example is Uganda, where the language of instruction is typically English, but rural schools may choose a dominant local language instead for the first three years of primary schooling (Masagazi Masaazi, Ssentanda, & Ngaka, 2018). Nevertheless, and due to the larger number of languages operating in rural contexts, rural students are unlikely to attend schools where the language of instruction is their own L1 (Orwenjo, 2012). Language policies in education sometimes assume that children living in an area where a particular language is widely spoken, automatically acquire that language, which can be far from reality (Nyarko Ansah, 2014).

In addition to the rural/urban divide, the impact of language policies on students' identity can be moderated by immigration status. Identity development is complex among immigrants, as their identities fluctuate depending on the context in which they function, and are shaped to meet the expectations of both host and home communities (Vandeyar & Catalano 2020; Vandeyar & Vandeyar 2011). In particular, proficiency in host language(s) is key in determining educational outcomes and prospects of second-generation migrants (Rassool, 2014), yet proficiency in home languages impacts their sense of belonging (Vandeyar & Catalano, 2020). The review highlighted that consciously or unconsciously, immigrants, including refugees, tend to prioritize host languages to gain acceptance in their new communities, ultimately experiencing language shift or home language loss (Dekoke, 2016; Reddick & Dryden-Peterson, 2021; Vandeyar & Catalano, 2020). Language shift or loss increases the challenges of properly communicating with families or communities of origin, impacting students' wellbeing (Reddick & Dryden-Peterson, 2021). Language choices in education can feel like a losing game for immigrants who may think of their multilingualism as a burden (Dekoke, 2016).

Beyond identity, teaching and language practices may affect immigrants and refugees' ability to learn in unique ways. For example, pedagogical resources as *translanguaging* – or flexible actions in which teachers aim to draw from students' full linguistic repertoires (Velasco & Garcia, 2014) – play out differently for immigrant students, as the languages used in that process may be languages that they do not know or understand (Vandeyar & Catalano, 2020; Vandeyar & Vandeyar, 2011). Or teachers in refugee camps may struggle with bilingual education and the degree of linguistic heterogeneity of their students, resorting to official languages (e.g., English) to educate students (Dryden-Peterson et al., 2019), affecting refugees' learning outcomes and sense of identity. Pedagogical and language choices in the classroom affect immigrants and refugees in ways that are neither obvious nor explicitly considered in bilingual education programs.

Language policies, furthermore, interact with some other socioeconomic background characteristics of students. For instance, the linguistic capital of students – particularly, the language(s) spoken at home – may mediate the reproduction of social class in educational outcomes (Nishioka & Durrani, 2018). This means that students who come from privileged backgrounds may increase their advantages through the alignment between their home

languages and the languages spoken at school. Or linguistic choices may affect girls and boys differently. For example, Penner Angrist (2018) suggests that other things being equal, using French as the language of instruction created additional hurdles to girls' school attendance in West and Central Africa.

5.1.2. Policy design challenges

The review pointed to three main problems related to the conception of bilingual education programs. First, policies are often not conceived from the bottom-up and can be detached from reality and overlook the dynamic nature of multilingualism in Sub-Saharan Africa. Language-in-education policies in Africa in many cases do not reflect the fluid bilingual identities of their inhabitants (Nyarko Ansah, 2014) and the inherent heteroglossia of the languages (Rassool, 2014), and remain "extensively bookish and Euro-centric" (Bokamba, 2011). Top-down language policies, usually assume that children can perform as one or multiple monolingual students, in any given language and form, which is inaccurate. In addition, there is some consensus that politicians may prioritize languages that people do not speak, such as English (Borg, 2015; Sibomana, 2020), and ignore languages that people do speak, like Bantu languages and Khoi San languages (Makoni & Makoni, 2015). As such, the aspirations set for bilingual education are unrealistic and misaligned with existing policies (Ekkehard, 2018).

Second, bilingual education policies, often do not integrate all the relevant stakeholders in their scope, disregarding the views of certain stakeholders such as teachers, principals (Lazdowski, 2015), or parents (Evans, 2018). Bilingual education policies also, usually exclude the experiences and needs of groups such as immigrants and migratory groups (Rassool, 2014; Vandeyar & Catalano, 2020). Overlooking the views of stakeholders and the experiences of sub-groups could create biases and deepen the existing learning gaps. Lastly, language-in-education policies have been slow to incorporate best practices regarding learning outcomes. Even though there is evidence that links teaching students in languages they do not understand to lower performance (Seid, 2016), in some countries such as Malawi children are taught in a foreign language as early as the first grade (Nishioka & Durrani, 2018). In addition, lack of proficiency in the language of instruction can lead to students losing interest and motivation to attend school (Babaci-Wilhite, 2013; Jepkosgei Cheruiyot & Kabellow, 2018). Altogether, bilingual education policies that are detached from the sociolinguistic reality experienced by students, are biased towards some groups, and overlook pedagogical considerations can be counterproductive.

5.1.3. Resource challenges

The delivery of bilingual education requires a steady and sufficient flow of targeted resources. Resource challenges refer to those that are directly linked to shortages or shortcomings in materials and labour. Many authors have stressed that schools and teachers lacked textbooks, syllabuses, and curricula to teach in the languages they were tasked with (e.g., Caesar, 2011; Ekkehard, 2018; Igboanusi, 2014; Nyarko Ansah, 2014; Heugh & Mulumba, 2013). The shortage of materials impedes the ability of teachers to implement the language policies in place, whether in foreign or local languages. The problem is particularly more severe for local languages, as most of the resources are available in English and/or other foreign languages and not in indigenous languages (Makoni & Makoni, 2015). The lack of materials in the languages of instruction would require teachers to translate materials from languages such

as English onto the languages in which they are supposed to teach (Nyarko Ansah, 2014). These materials may carry Eurocentric biases (Evans, 2018), making them culturally unsuitable. Moreover, the scarcity of educational material outside schools, such as literature, further restricts the literacy development of children in local languages (Ekkehard, 2018). Overall, the lack of materials constitutes direct and an indirect impediment to the literacy acquisition of children in certain languages, regardless of the policies in place.

The review pointed towards another major resource challenge: the lack of teachers who can implement bilingual education as intended. Many authors have mentioned severe shortages of trained and qualified teachers in Sub-Saharan Africa (Babaci-Wilhite, 2013; Ekkehard, 2018), who are proficient in the languages of instruction (Borg, 2015; Evans, 2018; Sibomana, 2020) and are specifically trained to teach in those languages (Caesar, 2011; Heugh & Mulumba, 2013). Critically, teachers are not trained on second language methodologies and are expected to teach a second language relying on the same first language methodologies (Eno, Ahad, & Shafat, 2019). The lack of trained teachers coupled with large rates of teacher turnover and attrition in the region (Pitsoe & Machaisa, 2012) pose immense challenges to the success of bilingual education methods which heavily rely on the teachers' delivery. For example, the lack of teacher workforce prepared to teach in local languages is in and of itself an impediment to change policies that favour foreign languages (Igboanusi, 2014). Better training and professional development are needed to address these shortcomings. It should also be noted that teachers in the region tend to face incredibly harsh conditions, including stress, lack of career advancement, and low salaries (Babaci-Wilhite, 2013; Pitsoe & Machaisa, 2012). In addition, they are in some cases promoted and valued on measures other than their performance (Eno, Ahad, & Shafat, 2019). Therefore, beyond professional development, teachers may lack incentives and motivation to acquire the competencies they would need to successfully support multilingual education.

5.1.4. Ideology and conceptual challenges

Much of the debate around bilingual education is driven by ideological differences between stakeholders. Policymakers espouse multiple reasons to justify the predominance of languages such as English, French, or Portuguese in African education systems, bilingual or not. For instance, English is thought of as a language that would support the economic transformation that certain African nations want to pursue (Borg, 2015). In certain countries, there is almost an unconscious and automatic association between English and social progress and economic advancement (Igboanusi, 2014). The prevalence of foreign languages in the African education system could also be explained by the fact that it does in fact benefit the elites (Bokamba, 2011; Eno, Ahad & Shafat, 2019; Makoni & Makoni, 2015). However, the preference for foreign languages is widely shared by parents, teachers, and students themselves. Parents across African countries extensively believe that it is not beneficial to teach students in languages that they already know (Igboanusi, 2014), that English is the language of science (Orwenjo, 2012), that languages such as English open up more opportunities for their children (Eno, Ahad, & Shafat, 2019; Lazdowski, 2015), and may even equate English proficiency to intelligence (Muaka, 2011). For these reasons, they often insist on using foreign languages in instruction (Evans, 2018). Similarly, teachers do not always have a high regard for local

languages (Evans, 2018), and sometimes consider mother tongue-based education to be detrimental to the development of more useful languages such as English (Muaka, 2011). Last, students themselves, particularly immigrants, may have preferences for languages such as English that they consider more advantageous (Dekoke, 2016). Overall, the strong support to foreign languages on the side of parents, teachers, and students is at odds with their negative impact on identity development. Although, groups such as parents may prefer foreign languages because they could simply be unaware of other alternatives and their consequences (Orwenjo, 2012), the literature suggests that the reasons behind this preference can be more convoluted, partly explained by what Bokamba (2011) coins as *Ukolonia*, or an ideological stance that makes African people think of African languages as inherently inferior.

The literature highlights deeper ways in which belief systems underlie debates about bilingual education. Linguistic terms are not neutral, and their socio-political meanings vary by context. Accordingly, bilingualism is not inherently progressive or conservative (Makoni & Makoni, 2015). Bilingualism in Africa can be thought of in terms of language policies in which one foreign language occupies a higher status than an indigenous one, or alternatively, refer to situations where two foreign languages are used in official matters (Ekkehard, 2018). For example, in the DRC, colonial bilingualism tended to focus on the relation between French and Dutch and in apartheid South Africa, bilingualism traditionally referred to the relation between English and Afrikaans, as opposed to the relations between foreign and indigenous languages (Ekkehard, 2018; Meeuwis, 2011). Mother tongue-based education, while promoted by international organizations, may not be considered beneficial by Africans, perhaps because of the racial undertones of such policies in the past (Jepkosgei Cheruiyot & Kabellow, 2018). In addition, there are substantial differences in how African multilingualism is perceived by different stakeholders. For instance, some African political leaders subscribe to the ideology of language as a challenge (Makoni & Makoni, 2015, Orwenjo, 2012, Babra Gora, 2017). It is also argued that one language, typically foreign, has the potential to unite people, which has been called an outright myth by many (e.g., Babra Gora, 2017; Meeuwis, 2011; Orwenjo, 2012). Such positions can threaten any model of bilingual education. These debates are further complicated by the fact that languages are not static phenomena, and many hybrid languages are widely used in Sub-Saharan Africa, such as Kenya's Sheng, a hybrid of English and Kiswahili (Rassool, 2014). Discussions around bilingual education should go beyond the pluralization of singularity, or else, bilingual education is set to fail (Makoni & Makoni, 2015).

Misconceptions around language exist at the pedagogical level as well. A commonly held misconception is that receiving education in a language is the same as learning that language (Babaci-Wilhite, 2013). A closer look at the literacy outcomes in countries with former or current English-only policies such as Gambia, Rwanda, or Malawi would quickly refute such hypothesis (e.g., see discussions and outcomes in Borg 2015; Nishioka & Durrani 2018; Nyarko Ansah, 2014; Sibomana, 2020; Igboanusi, 2014). Teaching children in the weakest of their languages will negatively impact their learning outcomes without doubt. There is also a lot of discussion around cost effectiveness of teaching in local languages. Answering that question requires highlighting the pedagogical benefits of learning in local languages. There is empirical evidence from Sub-Saharan Africa that using local languages in schools improves

educational outcomes and reading achievement across multiple languages (Ramachandran, 2017; Seid, 2016). Using indigenous languages in education may have other benefits such as better female literacy outcomes (Penner Angrist, 2018) and higher social cohesion (Rassool, 2014). Or as the Tanzanian model shows, indigenous languages such as Kiswahili are sufficient to express scientific concepts and support learning beyond early grades (Makoni & Makoni, 2015). Therefore, while bilingual and mother tongue education are thought of as expensive, they can be cost-efficient (Ramachandran, 2017). As argued by Orwenjo (2012), an uneducated society would be a lot more expensive.

Another important pedagogical misconception is that languages in education are used in a homogeneous and standard way. The literature highlights that in almost every context, teachers translanguage and code-switch to deliver content (Caesar, 2011; Ekkehard, 2018; Evans, 2018; Jepkosgei Cheruiyot & Kabellow, 2018; Reddick & Dryden-Peterson, 2021; Sibomana, 2020). Part of the reason has to do with teachers' or students' lack of proficiency in the language of instruction. Yet, translanguaging is used for other purposes as well, such as conflict control (Jepkosgei Cheruiyot & Kabellow, 2018). It is worth mentioning that only one study mentions the misconceptions embedded in the language of assessments. In particular, Gándara and Randall (2019) show the potential benefits of using translanguaging when assessing bilingual students in smaller scale contexts and highlight the need to move away from strict monolingual practices in that regard. Overall, bilingual education in Africa needs to acknowledge the ways in which teachers and students use language inside and outside the classroom. Furthermore, bilingual education needs to be based in integrated approaches to teach languages (Ekkehard, 2018).

5.1.5. Linguistic challenges

The reviewed articles identified a series of challenges for bilingual education stemming from the linguistic characteristics of the African languages themselves. The most salient idea was that multiple indigenous and local languages are not developed enough to support instruction and learning (Igboanusi, 2014; Masagazi Masaazi, Ssentanda, & Ngaka, 2018). In particular, the absence of appropriate standardized orthographies (Ekkehard, 2018; Sibomana, 2020) and the lack of efforts to intellectualize African languages (Ekkehard, 2018) were perceived as high barriers to functional bilingual education. Some of the articles call for the establishment of language academies or national conferences for orthographic development of languages (e.g., Masagazi Masaazi, Ssentanda, & Ngaka, 2018). However, in some countries such as Uganda, language academies previously established are not currently functional (Caesar, 2011). In addition, language academies and efforts to standardize languages are criticized for pushing monolingual strategies that establish arbitrary yet authoritative norms for singular languages, contradicting the dynamic and multilingual reality of the African context (Makoni & Makoni, 2015). Such complications are heightened by a long history of language fragmentation. Back in colonial times, missionaries invested in the development of indigenous languages, and did so in ways that separated closely related languages (Babaci-Wilhite, 2013; Heugh & Mulumba, 2013), amplifying or directly inventing ethnic and linguistic diversity (Ekkehard, 2018). Therefore, beyond the need to develop languages, multiple scholars identify the additional challenge of bringing languages closer together, so that orthographies are accessible to more people (Heugh & Mulumba, 2013) and

match spoken practices (Ekkehard, 2018). Bilingual education and African societies in general, could benefit from thoughtful harmonization of language(s).

5.1.6. Historical challenges

The literature review highlighted numerous historical challenges that pervade Sub-Saharan African bilingual education. A salient challenge is the complicated legacy of colonial history and the way it frames language-in-education policies. Even after independence, colonial languages prevail in most African states (Bokamba, 2011). Colonial policies placed indigenous languages at the periphery, while giving currency to colonial languages such as English or French (Rassool, 2014; Nishioka & Durrani, 2018), installing the notion that African languages lacked value (Sibomana, 2020). However, colonial language policies played out differently in former British, Belgian, French, and other colonies. Belgian, British, and German colonizers prioritized African languages in education, while French colonizers attempted to assimilate people into the French society and culture by promoting a monolingual French state (Bokamba, 2011; Buzasi, 2016). Importantly, the promotion of indigenous languages by colonizers did not follow social justice motives (Bokamba, 2011). Indeed, policies favouring local languages had strong racial undertones (Makoni & Makoni, 2015) and were based on the idea that Africans were ignorant and could not properly learn colonial languages (Meeuwis, 2011). The degree to which African languages were used officially and in education also had to do with the level of involvement of the Church in state affairs. Using indigenous languages for proselytization was more effective in attracting new converts (Bokamba, 2011), and therefore, missionaries preferred educating people in these languages. French and Portuguese colonies minimized the role of the church in education and state affairs, leading to the prioritization of colonial languages (Bokamba, 2011). Bottomline, none of the colonial language policies can genuinely be considered altruistic. Monolingual models of education that prioritize a foreign language, bilingual models of education that contemplate foreign and indigenous languages, and mother tongue-based education can all be associated with a convoluted past.

The prevalence of colonial languages is often seen as an indicator for lack of absolute freedom (Nyarko Ansah, 2014; Babaci-Wilhite, 2013). Some authors believe that former colonial powers maintain their influence in the region, partly through language (Ekkehard, 2018; Lazdowski, 2015). For instance, some countries may have incentives to keep using colonial languages in education to foster and enable economic relations and/or budgetary assistance (Lazdowski, 2015). Indeed, neoliberalism and modernization themselves have been interpreted as neo-colonial forces by some, which continue to force indigenous languages to the periphery (Babra Gora, 2017) and put pressure on African countries (Ekkehard, 2018). Therefore, the use of multiple and/or indigenous languages in modern African states may mirror a common historical concern that it may curtail development and growth opportunities (Meeuwis, 2011). Monolingual education policies that prioritize colonial languages still have support in the region. Interestingly, this support may be contingent upon the historical relationship between language and identity in each country. Buzasi (2016) shows that the number of languages spoken at one's home has a correlation with identifying primarily as a citizen of a given nation, as opposed to a member of an ethnic group. The relationship holds for non-French and non-British former colonies, like the DRC. In addition, Buzasi (2016) shows that the likelihood of being

able to communicate with a randomly selected individual in the country is positively associated with identifying with the nation in former French colonies, but not elsewhere. These results point at different interpretations of the link between language and identity depending on the colonial history. Moreover, the historical relations between language, religion, and state may be linked to other outcomes. For instance, former French colonies rank at the very bottom in African female literacy (Penner Angrist, 2018). Overall, people's experience of colonial language policies varies by context and bilingual policies need to be understood within each specific context.

5.1.7. Implementation challenges

The literature review also highlighted challenges related to the implementation of language in education policies. On the one hand, scholars referred to lack of political will among African leaders to promote bilingual education. In their view, discourses around language-in-education do not match political action (Babra Gora, 2017). For example, political leaders may not be willing to allocate the funds that successful implementation of bilingual education requires (Eno, Ahad, & Shafat, 2019). Or governments may not be willing to provide support to key initiatives that would facilitate bilingual and mother tongue-based education, such as language development led by language academies (Masagazi Masaazi, Ssentanda, & Ngaka, 2018; Babaci-Wilhite, 2013). In addition, even when political leaders invest in the establishment of institutions to develop or protect indigenous languages, they may not endow them with the authority to act upon their mandates (Makoni & Makoni, 2015). Overall, some scholars argue that after independence, African political leaders were clearer on what they wanted to leave behind rather than what they wanted to change, and their eagerness to transform the status quo quickly faded away (Bokamba, 2011; Makhudu, 2017).

Bottomline, the reviewed articles suggest that the lack of success of bilingual education can be partly attributed to failure at the implementation level. In rural areas, despite policies that foster the use of local languages in education, students are rarely taught in their mother tongues (Muaka, 2011). Evans (2018) attributes the problem to a disjuncture between policy and support provided to teachers and other school-level actors. Pedagogical support matters, because teachers need tailored and well-thought-out training and materials to deliver education in local languages. Igboanusi (2014) highlights that African education systems face other challenges, such as poor infrastructure and lack of financial means, which can make the delivery of quality education difficult, irrespective of the language policy. In turn, Makhudu (2017) attributes implementation failures to poor organizational and leadership skills of those running the programs (Makhudu, 2017). And Nyarko Ansah (2014) states that implementation failure stems directly from poor policy design. Regardless of the motifs, there is latent belief that bilingual education also fails at the implementation level (Masagazi Masaazi, Ssentanda, & Ngaka, 2018). Successful models of bilingual education need to address the local and intermediate levels and acknowledge what happens at the school and/or regional level, to ensure a smoother and more consistent delivery.

6.Results

6.1.Research question 2: Representation of bilingual literacy

The assessments examined in this study were designed to evaluate literacy acquisition after each trimester of instruction. The assessments were neither too easy nor too difficult for students. The tasks were different between languages, because grade 3 students have higher levels of literacy in National Languages than French, and their language development is not the same across these languages. Tasks 1 and 5 measured essential skills in French, Task 2 measured essential skills in National Languages, and Task 6 measured essential skills across languages. Task 6 measured competency across languages (bilingual skills) and metalinguistic skills, and was used to examine the transference and development of writing skills for both languages.

Figure 7 Sample Items

TACHE 1 : MOTS FREQUENTS

Français : Cette tache a deux parties. Pour la première partie, je vais te montrer quelques mots que tu devras lire à haute voix. Par exemple, si je te montre le mot suivant (montrer le mot « elle ») peux-tu le lire à haute voix ? (Attendez jusqu'à ce que l'élève dise « elle ». Si l'élève répond correctement, dites « correct ». Si l'élève ne répond pas correctement, expliquez-lui que la réponse correcte était « elle »). Je vais maintenant te demander de faire la même chose pour un groupe de mots. Tu dois pointer du doigt le mot que tu lis. Tu vas lire de gauche à droite et de haut en bas. Es-tu prêt ? Commençons.

| | | | |
|---------|-------|---------|--------|
| à | papa | bic | car |
| ceci | sur | cette | un |
| maman | après | élève | même |
| pas | puis | bien | chaque |
| bientôt | enfin | dessous | eau |

(a) Task 1 Frequent Words

TACHE 2: LECTURE EN LANGUE NATIONALE

Français : Maintenant, je vais te montrer une histoire et je voudrais que tu la lises à haute voix mot par mot. Tu peux pointer du doigt le mot que tu vas lire. Si tu ne sais pas comment lire un mot tu peux passer au mot suivant. Cet exercice n'est pas chronométré mais parfois il est possible que je te demande de passer au mot suivant. Es-tu prêt ? Commençons

Cisalu cinene

Mamu, Kabedi ne Kalala bavua baya mu cisalu.
Mu cisalu movua bantu bituilangana, kakuyi kua kupicila.
Bapanyishi bela mbila bua kubikila basumbi.
Mamu ekusumbila Kabedi nkanzu, Kalala bisabata.
Mu cisalu mudi mitoyi ya bungi.

(b) Task 2 Ciluba ORF

TACHE 5 : DÉCODAGE**Partie I :**

Français : Cette tache a quatre parties. Pour les deux premières parties, je vais te montrer sept mots et tu devras faire des groupes avec ces mots. Pour la première partie, tu devras grouper les mots à partir du son que tu entends au début des mots. Par exemple, si je te montre les mots suivants (Montrez la première liste de mots : table, cahier...etc.) comment pouvons-nous grouper les mots qui ont le même son au début ? (Attendez jusqu'à ce que l'élève essaie de grouper les mots. Si l'élève n'essaie pas de grouper les mots, montrez-lui comment le faire et expliquez-lui pourquoi vous avez mis certains mots ensemble. Si l'élève essaie de grouper les mots et il/elle le fait correctement, dites « correct ». Si non, montrez à l'élève la réponse correcte et expliquez-lui quels étaient les erreurs). Je vais maintenant te demander de faire la même chose encore une fois. Es-tu prêt ? Commençons.

Liste 1 (pratique) :

Taille
Table
Cahier
Quatre
Kola
Tasses
Cuisine

Partie II :

Français : Pour la deuxième partie, tu devras grouper les mots à partir du son que tu entends à la fin des mots. Par exemple, si je te montre les mots suivants (Montrez la liste de mots : dame, oeuf, etc.). Comment pouvons-nous grouper les mots qui ont le même son à la fin ? (Attendez jusqu'à ce que l'élève essaie de grouper les mots. Si l'élève n'essaie pas de grouper les mots, montrez-lui comment le faire et expliquez-lui pourquoi vous avez mis certains mots ensemble. Si l'élève essaie de grouper les mots et il/elle le fait correctement, dites « correct ». Si non, montrez à l'élève la réponse correcte et expliquez-lui quels étaient les erreurs). Je vais maintenant te demander de faire la même chose encore une fois. Es-tu prêt ? Commençons.

Liste 2 (pratique) :

Dame
Oeuf
Lame
Neuf
Comme
Girafe
Pomme

(c) Task 5 Decoding Parts I and II**Partie III :**

Français : Pour la troisième partie, je vais lire trois mots et tu devras identifier quelle paire de mots a le même son au début. Par exemple, si je lis « exemple – tableau – exercice » quels sont les deux mots qui ont le même son au début ? (Attendez jusqu'à ce que l'élève essaie de répondre. Si l'élève répond correctement, dites « correct ». Si non, expliquez-lui que la réponse correcte était « exemple » et « exercice » parce que les deux mots commencent avec le son « gz »). Je vais maintenant te demander de faire la même chose avec quatre groupes de mots. Es-tu prêt ? Commençons.

Mots Partie III – Pratique :

1. Exemple Tableau Exercice

Partie IV :

Français : Pour la quatrième partie, je vais lire trois mots et tu devras identifier quelle paire de mots a le même son à la fin. Par exemple, si je lis « luxe – photo – axe » quels sont les deux mots qui ont le même son à la fin ? (Attendez jusqu'à ce que l'élève essaie de répondre. Si l'élève répond correctement, dites « correct ». Si non, expliquez-lui que la réponse correcte était « luxe » et « axe » parce que les deux mots finissent avec le son « ks »). Je vais maintenant te demander de faire la même chose avec quatre groupes de mots. Es-tu prêt ? Commençons.

Mots Partie IV – Pratique :

1. Luxe Photo Axe

(d) Task 5 Decoding Parts III and IV**TACHE 6 : PRODUCTION ECRITE****Partie I :**

Français : Cette tâche a trois parties. Pour la première partie, tu devras écrire différentes phrases en français. Voici quelques feuilles et un stylo pour compléter la tâche. Je vais lire chaque phrase deux fois. Je vais t'indiquer quand j'ai fini en disant « vas-y, écris la phrase ». Après, tu vas écrire la phrase que j'ai dite une seule fois. D'accord ? Par exemple, je vais lire une phrase à haute-voix (Lisez « C'est une femme. » deux fois et dites « vas-y, écris la phrase ». Attendez jusqu'à l'élève écrive « C'est une femme. » Si l'élève écrit la phrase correctement, dites « correct ». Si l'élève n'écrit pas correctement la phrase, montrez-lui comment écrire « C'est une femme. » Pratiquez une deuxième fois s'il le faut). Je vais maintenant te demander la même chose pour des autres phrases. Es-tu prêt ? Commençons.

Sample Item :

1. Sur la table il y a quatre kolas.

Partie II :

Français : Pour la deuxième partie, tu devras écrire différentes phrases en lingala. Tu peux utiliser les mêmes feuilles et le stylo pour cette tâche. Je vais lire chaque phrase en lingala deux fois. Je vais t'indiquer quand j'ai fini en disant « vas-y, écris la phrase ». Après, tu vas écrire la phrase que j'ai dite une seule fois. D'accord ? Par exemple, je vais lire une phrase à haut-voix (Lisez « Masanga manso makutanaka. » deux fois et après dites « vas-y, écris la phrase ». Attendez jusqu'à l'élève écrive « Masanga manso makutanaka. » Si l'élève écrit la phrase correctement, dites « correct ». Si l'élève n'écrit pas correctement la phrase, montrez-lui comment écrire « Masanga manso makutanaka. » Pratiquez une deuxième fois s'il le faut). Je vais maintenant te demander la même chose pour d'autres phrases. Es-tu prêt ? Commençons.

Sample Item :

1. Tata ateki nyama.

(e) Task 6 Bilingual Writing

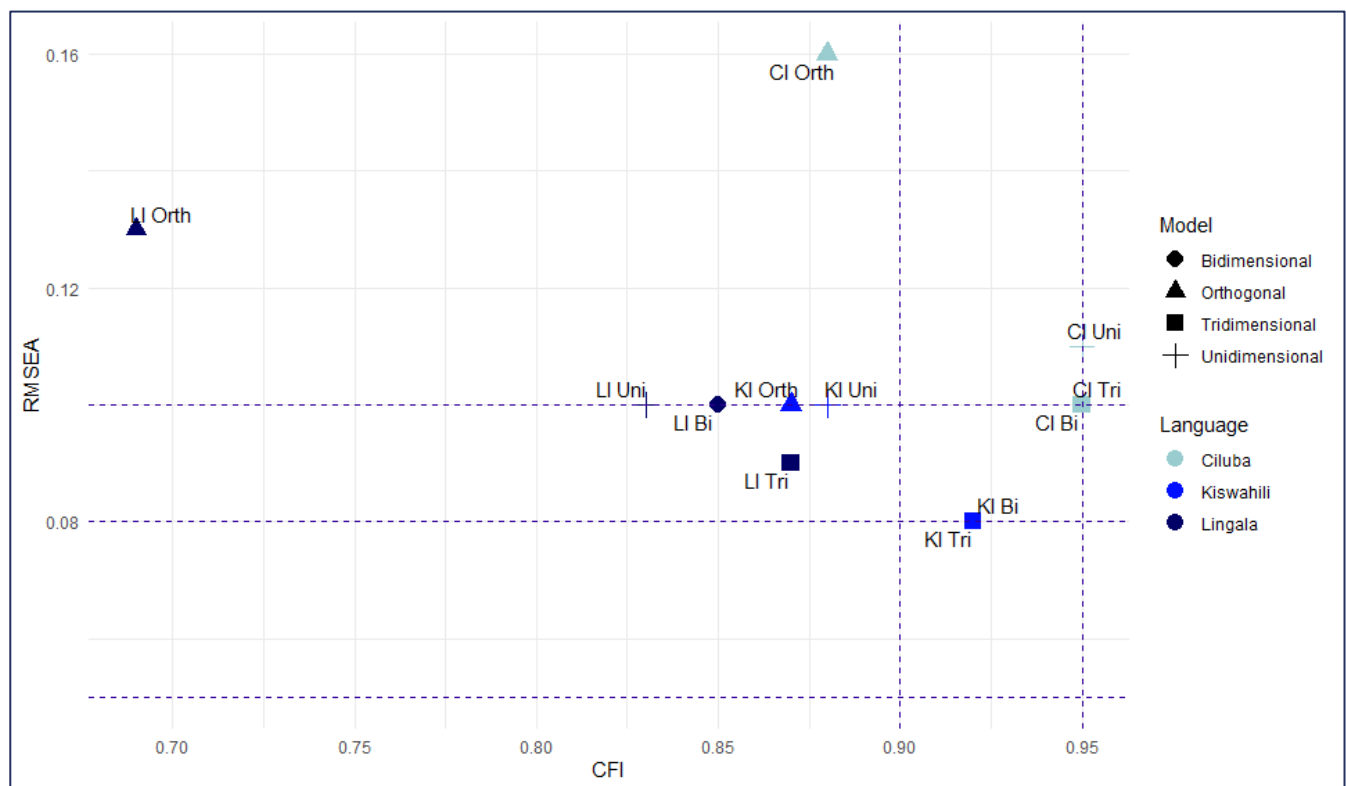
Table 10 shows the performance of students by task and provincial group. As shown, in average, scores indicate moderate difficulty, and range between 43% and 60% for the entire test. While scores are presented as percent correct, tasks other than French are not comparable across language groups, as they had not been equated.

Table 10 Performance by Task and Provincial Group (Percent Correct)

| | Task 1 - Automated Reading (French) | Task 2 - Oral Reading ¹⁶ (Normalized, National Language) | Task 5 - Grouping Words (French) | Task 6 - Writing (Bilingual) | Total |
|---------------------|-------------------------------------|---|----------------------------------|------------------------------|-------|
| Ciluba Provinces | 38% | 61% | 49% | 13% | 55% |
| Lingala Provinces | 45% | 65% | 59% | 10% | 60% |
| Kiswahili Provinces | 31% | 46% | 46% | 6% | 43% |

Four models were fit on the dataset: unidimensional, bidimensional, tridimensional, and orthogonal. Models used parcels of 4–9 items each, depending on the task. The models were evaluated primarily on two grounds: fit indices and the characteristics of the estimated parameters. Based on fit indices, the best model was different depending on the provincial group. For the Lingala provinces, the best performing model was the tridimensional model. For the rest of the groups, both the bidimensional and tridimensional models fit the data well. Across all provincial groups, the worst performing model was the orthogonal model (see Figure 8).

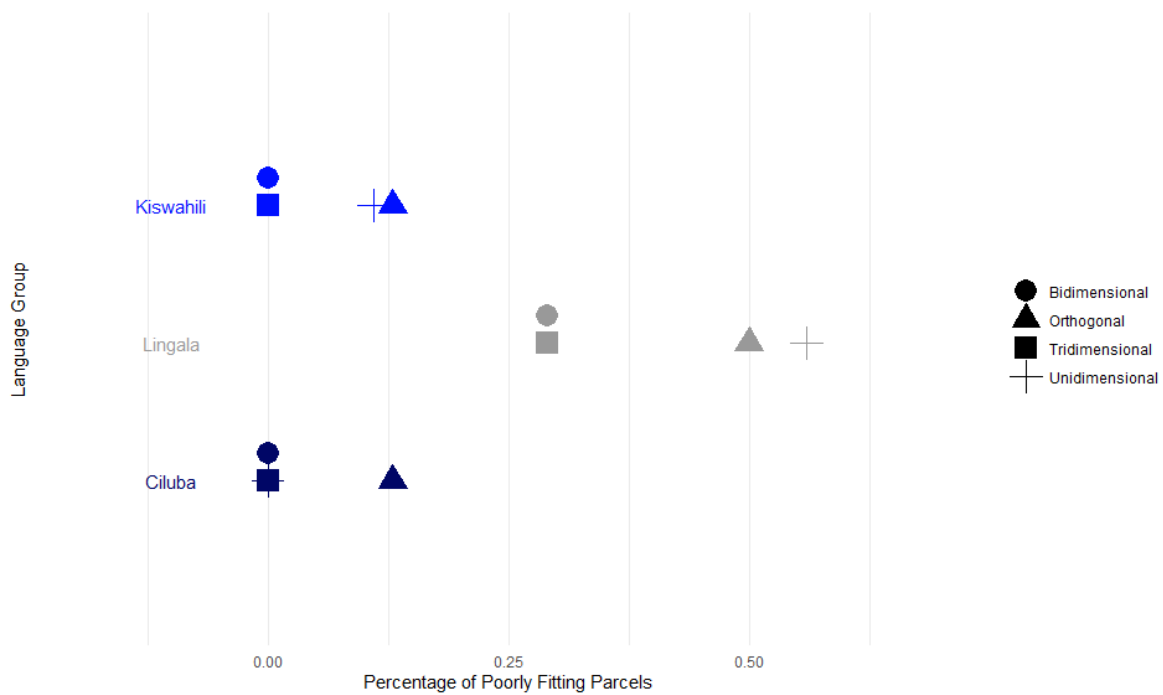
Figure 8 RMSEA and CFI for Models



¹⁶ Untimed. No fluency measures reported.

To further explore model performance, factor loadings for the different parcels of items were analysed. A parcel was considered poorly fitting when its estimated factor loading was lower than 0.4, negative, or statistically non-significant. Figure 9 shows the percentage of poorly fitting parcels by language group and model. There were few differences in terms of misfitting parcels between the models. Overall, the Kiswahili and Ciluba models showed less misfitting parcels than the Lingala models. The unidimensional and orthogonal models showed more problems in terms of fit than the rest of the models. The bidimensional and tridimensional models performed equally well for all language groups (see Figure 9)

Figure 9 Percentage of Misfit for Different Models



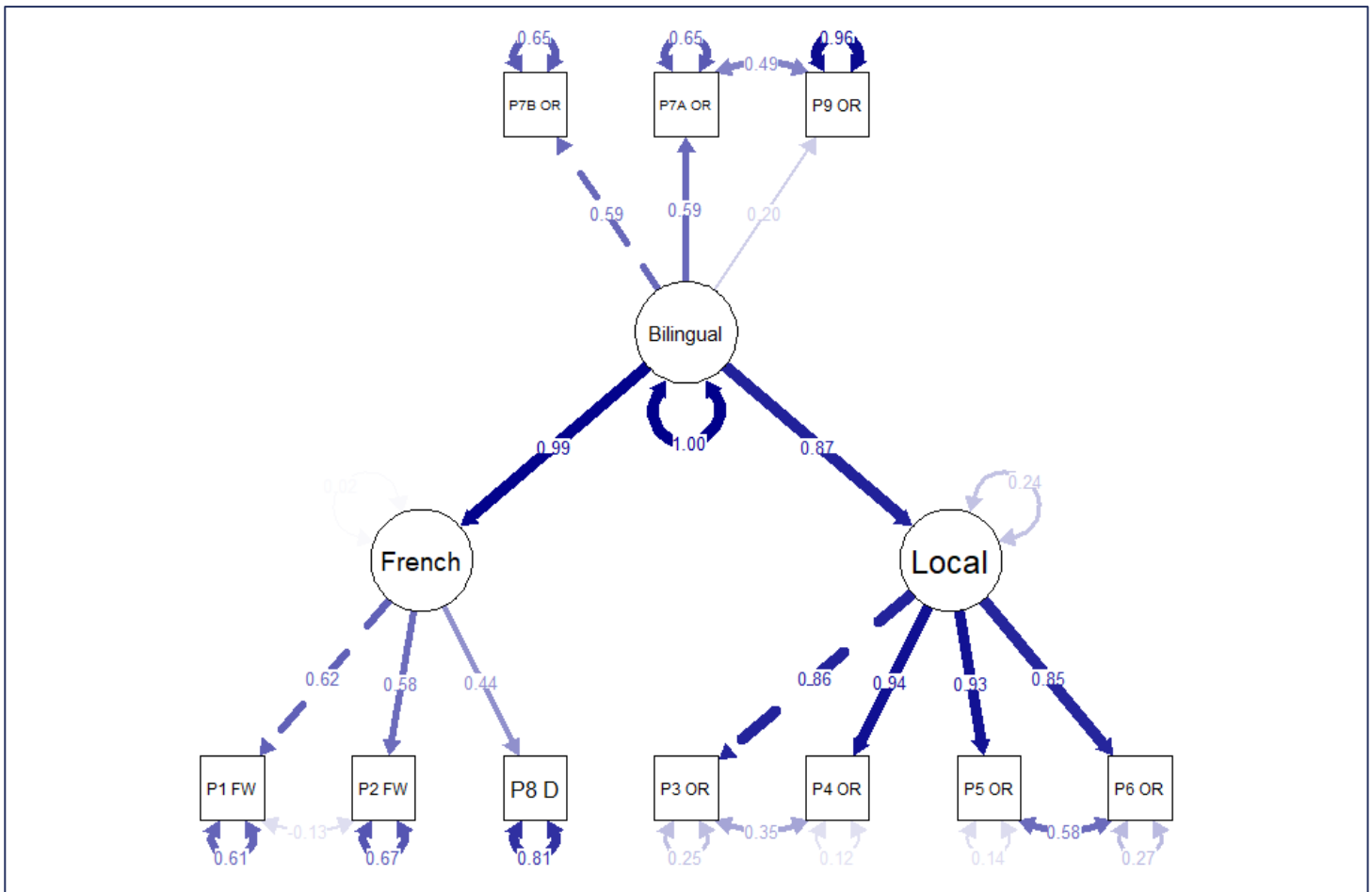
The previous analyses showed that both the bidimensional and tridimensional models performed acceptably well. A way to decide between the two models is through the χ^2 test. A p-value smaller than 0.05 on the χ^2 test would indicate that the tridimensional model fits the data better. Based on the results shown in Table 11, the bidimensional model is the best performing model across the Ciluba and Kiswahili groups, yet the tridimensional model is the best performing in the Lingala group.

Table 11 χ^2 test for bidimensional and tridimensional models

| | χ^2 | df | χ^2_{diff} | Δ df | P-Value |
|------------------|----------|----|-----------------|-------------|---------|
| Ciluba | | | | | |
| Bidimensional | 155 | 29 | | | |
| Tridimensional | 154 | 28 | 1 | 1 | 0.26 |
| Lingala | | | | | |
| Bidimensional | 105 | 29 | | | |
| Tridimensional | 93 | 28 | 13 | 1 | 0.00 |
| Kiswahili | | | | | |
| Bidimensional | 204 | 29 | | | |

The bidimensional model made a strong assumption about the relationship between bilingual skills and literacy in French and National Languages. However, one can conclude that the most appropriate models were those that acknowledged a complex relationship between languages and included a bilingual factor. The literacy of bilingual students is not a unidimensional construct. And the performance across languages should not be thought of as independent. Taking all the evidence together, the bidimensional models were slightly superior to the tridimensional models, except for the Lingala provincial group. Best performing models are displayed in Figure 10–12.

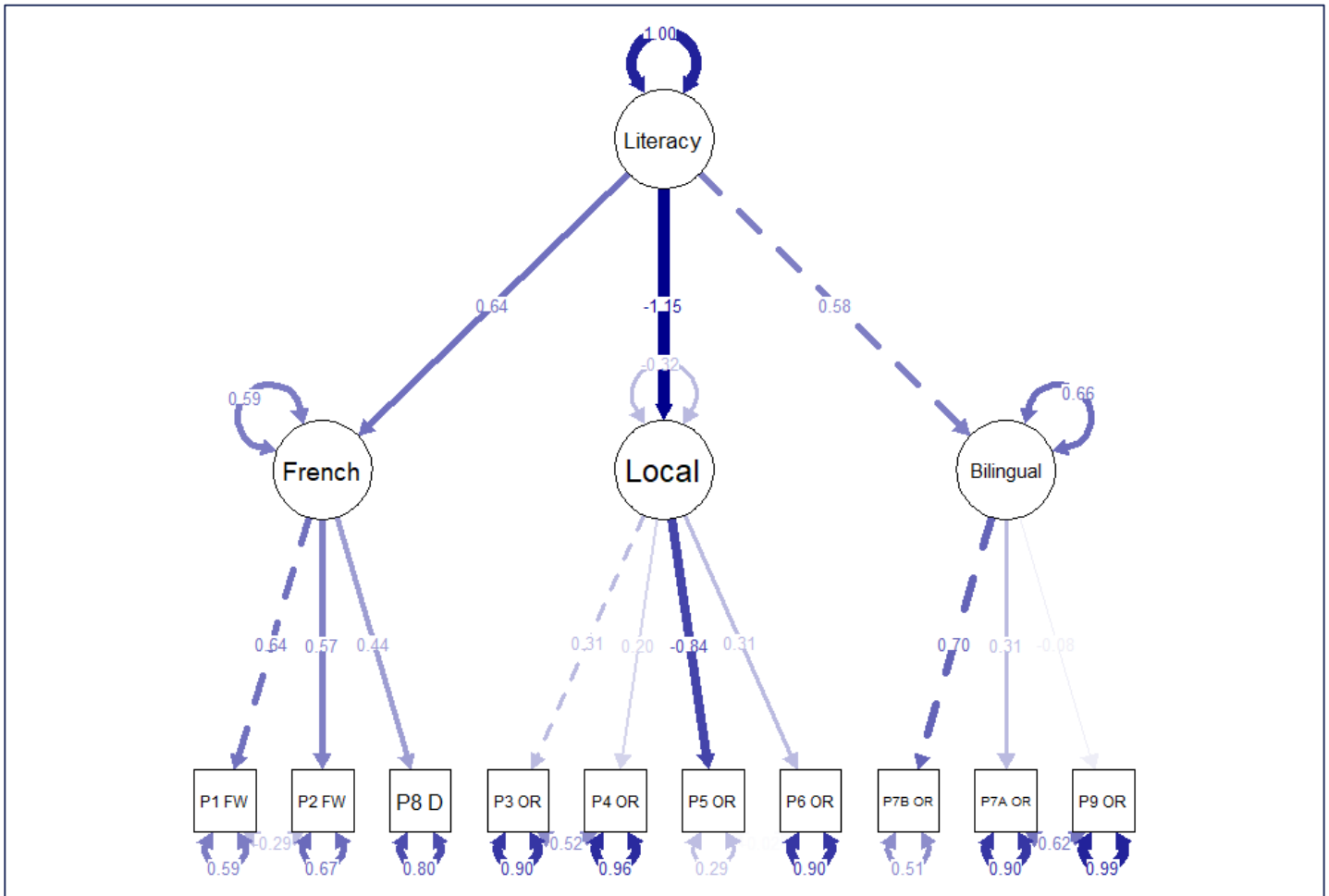
Figure 10 Best Model, Ciluba



Where:

- P1 FW – Parcel on frequent words
- P2 FW – Parcel on frequent words
- P8 D– Parcel on decoding
- P3 OR – P6 OR – Parcels on Oral Reading
- P7A W, P7B W – Parcels on Writing, National Languages and French
- P9 W– Parcels on Writing, Conventions

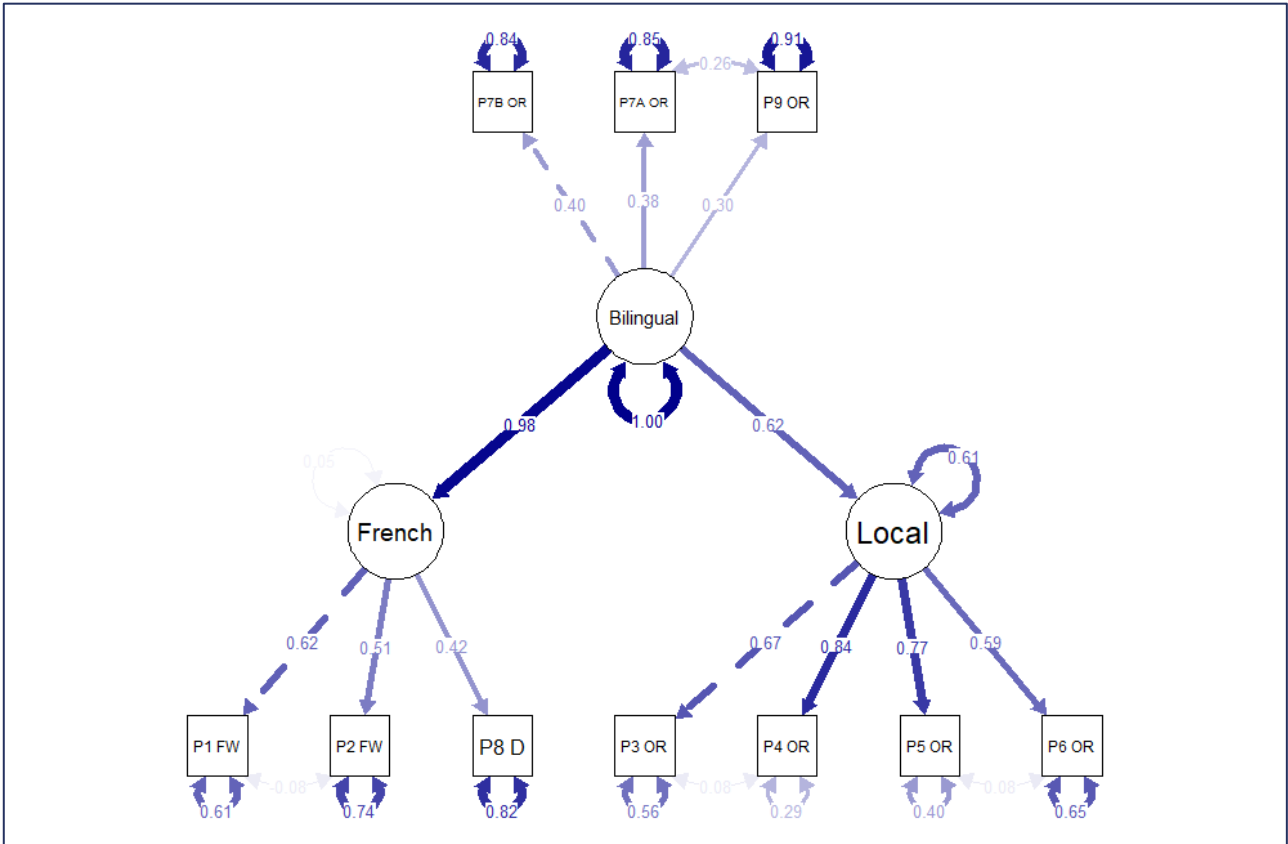
Figure 11 Best Model, Lingala



Where:

- P1 FW – Parcel on frequent words
- P2 FW – Parcel on frequent words
- P8 D– Parcel on decoding
- P3 OR – P6 OR – Parcels on Oral Reading
- P7A W, P7B W – Parcels on Writing, National Languages and French
- P9 W– Parcels on Writing, Conventions

Figure 12 Best Model, Kiswahili



Where:

- P1 FW – Parcel on frequent words
- P2 FW – Parcel on frequent words
- P8 D– Parcel on decoding
- P3 OR – P6 OR – Parcels on Oral Reading
- P7A W, P7B W – Parcels on Writing, National Languages and French
- P9 W– Parcels on Writing, Conventions

6.2. Research question 3: Equivalence of bilingual assessments

A different lens for evaluating bilingual assessments is provided through invariance. Invariance analyses were conducted on groups of students using two criteria: a) linguistic diversity experienced at home, and b) whether students spoke a minority language at home. The analyses were conducted using the best fitting models only: bidimensional model in the Ciluba and Kiswahili groups, and tridimensional model in the Lingala group. The main outcomes of the analysis are presented in Figure 13. Overall, the results indicated that none of the assessments

reached scalar levels of equivalence across subgroups¹⁷. The lack of scalar invariance prevents meaningful comparison between the means of the different subgroups, and disaggregating proficiency levels by linguistic background would not be appropriate. Equivalence was higher for the minority languages groups, for the Lingala and Kiswahili assessments. This could be explained by the fact that the likelihood of speaking minority languages at home is considerably lower in the Ciluba provinces (see Figure 5), and such a small group could be different in other ways (e.g., considerably poorer).

Figure 13 Invariance Analyses on the Best Performing Model

| | Linguistic Diversity | | | Minority Languages | | |
|------------|----------------------|-----------------------|-----------|--------------------|---------|-----------|
| | Ciluba | Lingala ¹⁸ | Kiswahili | Ciluba | Lingala | Kiswahili |
| Scalar | | | | | | |
| Metric | | | | | | |
| Configural | | | | | | |

Table 12 shows the sources of error for each solution. An invariance analysis places equality restrictions for the SEM solutions between different groups. Invariant models are such that if some restrictions of equality are lifted, the model fit improves considerably. Problematic parameters are highlighted in blue. As it can be seen in Table 12, sources of misfit are specific to each provincial group and linguistic disaggregation. There is not a clear trend as to which set of items were more problematic for the entire sample. Notably, the impact of bilingual skills on French is structurally different for linguistic minorities and other students in the Ciluba provinces. In addition, Table 12 shows that bilingual parcels were the best fitting across all solutions.

Table 12 P-values for improvement in fit if restriction lifted, by parameter

| | Linguistic Diversity | | | Minority Languages | | |
|-------------------------------------|----------------------|---------|-----------|--------------------|---------|-----------|
| | Ciluba | Lingala | Kiswahili | Ciluba | Lingala | Kiswahili |
| Factor Loading | | | | | | |
| Parcel 1 - Frequent French Words | | | | | | |
| Parcel 2 - Frequent French Words | 0.06 | 0.89 | 0.00 | 0.65 | 0.47 | 0.94 |
| Parcel 3 - ORF in National Language | | | | | | |
| Parcel 4 - ORF in National Language | 0.37 | 0.08 | 0.81 | 0.80 | 0.63 | 0.35 |
| Parcel 5 - ORF in National Language | 0.00 | 0.37 | 0.78 | 0.55 | 0.60 | 0.33 |
| Parcel 6 - ORF in National Language | 0.00 | 0.00 | 0.34 | 0.13 | 0.59 | 0.48 |
| Parcel 7A - Writing words - FR | | | | | | |
| Parcel 7B - Writing words - NL | 0.07 | 0.08 | 0.82 | 0.48 | 0.36 | 0.66 |
| Parcel 8 - Decoding in French | 0.82 | 0.01 | 0.24 | 0.48 | 0.84 | 0.96 |

¹⁷ For full statistical details, refer to the appendices

¹⁸ Approximated by bidimensional model

| | | | | | | |
|-------------------------------------|------|------|------|------|------|------|
| Parcel 9 - Written conventions | 0.40 | 0.53 | 0.15 | 0.57 | 0.79 | 0.13 |
| Latent Variables | | | | | | |
| Bilingual literacy | | | | | | |
| National languages literacy | 0.87 | 0.34 | 0.14 | 0.46 | 0.70 | 0.91 |
| French literacy | 0.54 | 0.35 | 0.20 | 0.03 | 0.68 | 0.51 |
| Intercept | | | | | | |
| Parcel 1 - Frequent French Words | | | | | 0.00 | 0.90 |
| Parcel 2 - Frequent French Words | | | | | 0.01 | 0.41 |
| Parcel 3 - ORF in National Language | | | | | 0.31 | 0.28 |
| Parcel 4 - ORF in National Language | | | | | 0.22 | 0.62 |
| Parcel 5 - ORF in National Language | | | | | 0.34 | 0.00 |
| Parcel 6 - ORF in National Language | | | | | 0.05 | 0.16 |
| Parcel 7A - Writing words - FR | | | | | 0.55 | 0.00 |
| Parcel 7B - Writing words - NL | | | | | 0.14 | 0.28 |
| Parcel 8 - Decoding in French | | | | | 0.18 | 0.04 |
| Parcel 9 - Written conventions | | | | | 0.06 | 0.19 |

Importantly, invariance does not mean fairness. For example, these assessments may be invariant for students of different linguistic backgrounds yet students from linguistic minorities may be underperforming, nevertheless. In certain regions, students from linguistic minority groups may have very low levels of proficiency in both French and the corresponding National Language (e.g., see Gibson, 2019). Bilingual assessments that do not measure minority languages could reach high levels of invariance and be as problematic as monolingual assessments in failing to capture students' proficiency.

7. Conclusions

This study examined literature and data on the topic of bilingual education and bilingual assessments in Sub-Saharan Africa. The in-depth review of challenges showed the enormous complexity faced by bilingual education programs. Challenges are numerous and immense, at times fall outside the scope of educational policy, and there is not always consensus on how to tackle them. For clarity, in this study, challenges were classified into seven categories. Based on this classification, the most frequent challenges corresponded to sociocultural, resource constraints, and policy design problems. Such prominence emphasizes the intricate relationship between successful bilingual education programs and the social, cultural, financial, and political context in which they are embedded. The review also showed that several challenges faced by bilingual education are related to structural conditions such as poverty, migration, or ethnolinguistic diversity, which are not in control of education policy makers. Few challenges fell clearly within the realm of educational policy, such as policy design. Overall, these findings suggest that improving bilingual education requires a comprehensive response that is aligned to other developmental goals put forth by countries.

Notably, the review of challenges of focus in Sub-Saharan Africa does not highlight themes that are prominent in today's discussions around educational policy. For instance, fairly absent is the topic of students with disabilities, which can introduce a whole new layer of complication to the discussion on bilingual education and assessments. Deaf and hard of hearing students are bilingual students, who use sign language and printed language in education. Yet none of the reviewed articles covered the topic of bilingual education from the perspective of students with disabilities. In addition, the reviewed articles rarely mentioned the opportunities that information and communications technology (ICT) bring to bilingual education policy and practice. ICT and other forms of technological advancement could support efforts to develop languages, create and disseminate teaching material, engage teachers in different forms of peer support and/or professional development, and expedite policy implementation, to name a few. Only one of the documents referred to the topic of assessing bilingual students, which is the core theme of this article. These gaps suggest that in order to increase their impact, researchers need to expand classical lines of inquiry in bilingual education by engaging in topics that investigate the immediate future of educational practice.

A different set of conclusions emerges in relation to the analysis of bilingual assessment data in the DRC. The structural analysis indicated that the literacy of bilingual students is best represented by a complex structure. In particular, bilingual data is best modelled by including a Bilingual Literacy factor, directly accounting for the responses of bilingual items. In the Ciluba and Kiswahili provinces, the Bilingual factor directly influenced the French and National Language factors, in accordance with programmatic assumptions about transference of skills between languages and their influence on individual languages. In the Lingala provinces, the best fitting model was the trilingual model, where bilingual skills did not directly impact Lingala and French skills. The underperformance of orthogonal models confirms that treating skills in each language as independent parameters is not appropriate. This result can be further justified considering that grade 3 students are learning different languages simultaneously and may transfer skills across those languages. The lack of support for unidimensional models shows that there are meaningful differences between the languages that need to be captured in any evaluation effort. This is not unexpected, given the major linguistic differences between the French and each National Language. Had students been tested on linguistically closer languages, the unidimensional model could have emerged as the better model. Overall, the results support the interdependence hypothesis and the need to measure skills across all languages in integrated ways.

The study also examined the equivalence of bilingual assessments. The results show that bilingual assessments are not necessarily equivalent across groups of students with different linguistic backgrounds. The assessments did not reach scalar levels of invariance, and in most cases, not even metric levels of invariance. Score equivalence was higher when grouping students by whether they spoke minority languages, yet it was impossible to determine the roots of equivalence based on the data alone. The results highlight important differences in the sources of misfit by provincial group and language disaggregation. In other words, the reasons why invariance was not met were different for the Ciluba, Lingala, and Kiswahili provinces, as well as for the two classifications of students. The

inconsistency suggests that early grades' bilingual assessments may hold different psychometric properties and structural relations across language groups. Moreover, none of the French, Ciluba, Lingala, or Kiswahili tasks alone guaranteed equivalence for students of different linguistic backgrounds.

Together, these results have implications for large-scale monitoring and evaluation. Policymakers recognize that bilingual students know multiple languages and are multiliterate, yet they are often asked to report single indicators of performance. Such is the case with SDG indicator 4.1.1. In the case of literacy, policymakers may decide to report proficiency in the country's official language of instruction only. The results of this study suggest that doing so would be problematic, as bilingual students' literacy is a complex construct and cannot be viewed through the lens of a single language. Some may argue that if proficiencies across languages are correlated, measuring a single language would suffice. The results of this study suggest otherwise. The underperformance of unidimensional models shows that proficiency in French and National Languages are not sufficiently correlated to justify taking one measure as a proxy for the other. Moreover, taking the proficiency in one language as a proxy for the other would necessitate more than just a high correlation between the scores but also some form of linking between them. Alternatively, policymakers may examine students' full linguistic repertoire and report a single proficiency score which accounts for their performance in multiple languages. In that case, policymakers would need to assess students on what they are expected to know in each language as opposed to using the exact same proficiency measure across languages. For example, students in this study took a task that measured their Oral Reading skills in National Languages but not in French. Having assessed their Oral Reading in French would have resulted in very low scores, and those scores would have had little to do with their proficiency and a lot to do with their developmental trajectories. Early grade students who are enrolled in transitional bilingual programs cannot be thought of as balanced bilinguals or as double monolinguals. In turn, this study shows that literacy can be meaningfully assessed using bilingual assessments, even at large scales. Once again, the underperformance of orthogonal models shows that better insight can be provided when students' languages are assessed in integrated manners, measuring skills that are relevant to the curriculum. The results of this study suggest that capturing skills across languages is adequate for early grade students who are learning multiple languages in uneven ways. Last, the lack of invariance between linguistic subgroups implies that high-stakes use of these assessments can be problematic. To use bilingual assessments for monitoring and large-scale evaluation, considerably more work is required around invariance and score interpretation. The sources of misfit are not consistent across provincial and language groups and fixing the problem would entail different solutions by context. These results suggest that adequate bilingual assessments may be context specific. Therefore, bilingual assessments may be better suited to support context-specific functions such as local monitoring or pedagogy, other than global monitoring and other aggregated uses of data.

Bilingual assessments remain promising and monolingual assessments remain problematic. More research is needed to understand the capacity of bilingual assessments to support pedagogy, monitoring, and evaluation. This study provides evidence from one context, and it is unclear whether the outcomes can be extended to dissimilar circumstances. Future studies should evaluate bilingual assessments under new scenarios, examining other

combinations of language(s) and skills. This study did not discuss in depth the complexities of how to choose models or conceptualize biliteracy. For example, in this study, languages were conceived as finite and distinguishable sets, which is only one way of thinking about languages. Further research should explore alternative conceptualizations of biliteracy and the pertinence of bilingual assessments to support those conceptualizations. Also, more research is needed to uncover the utility of integrated bilingual assessments in supporting pedagogy. If bilingual assessments do in fact provide a more comprehensive image of students' proficiency, they should be considered best suited to support instruction. In addition, and particularly relevant to large-scale monitoring, further efforts around bilingual assessments should ensure score equivalence for all subgroups of students. Extensive work is needed to understand the peculiarities of the tested language(s), the examinees, and the ways in which they interact. If scalar levels of invariance are not met, bilingual assessments cannot support meaningful comparisons between subgroups, or the inferences and decisions that would normally arise from those comparisons. To conclude, bilingual assessments deserve more attention from research and practice, and ought to be seriously considered in current efforts to improve bilingual education.

POLICY RECOMMENDATIONS

A key policy recommendation following the literature review is to **support local language academies to gather, produce, and disseminate in-depth and detailed understanding of the linguistic characteristics of each language used in education**. A higher level of knowledge of the languages used in bilingual education is critical to improving these programs. In addition, countries **should invest as early as possible in sociolinguistic mapping exercises**, to understand the linguistic makeup of students, teachers, and the ways in which languages are used across different regions. In collaboration with literacy and education specialists, these two recommendations should be conducive to the identification of developmental trajectories followed by typical students across regions, as they learn how to read and write in multiple languages. Studying those trajectories can better inform key decisions within bilingual programs, including the development of curricula, teaching and learning materials, trainings, and assessments of literacy, including bilingual assessments. To make such efforts cost-efficient, governments are encouraged to collaborate with other countries working on similar challenges. Language academies, sociolinguistic mapping exercises and the development of materials and assessments could be co-funded by countries working with the same languages. Due to high costs and the dynamic nature of languages, continuous collaboration between countries and experts is critical for the sustainability and relevance of these initiatives.

Furthermore, the results of this study show that bilinguals' literacy is a complex construct, and that it would be suboptimal to measure their languages independently. This study also highlights the value of using bilingual and metalinguistic tasks in literacy assessments. Altogether, these results have implications for classroom and small-scale assessments of literacy in the early grades. It is more appropriate and more efficient to examine students' literacy in an integrated manner, and **schools adopting bilingual education should incorporate bilingual assessments into their formative assessment practice**. Bilingual assessments may provide more nuanced formative data than independent monolingual assessments. The results also have implications for large-scale evaluations. Notably, early grade students whose language(s) develop in uneven ways cannot demonstrate their full proficiency by means of monolingual assessments. Prioritizing one language in large-scale assessments may be practical but it is not fully appropriate from a validity standpoint. If large-scale monitoring is concerned with literacy, as opposed to literacy in a specific language, then **large-scale assessments should be expanded to include more languages and to capture bilingual and metalinguistic skills**. Collecting large-scale bilingual data will enable richer conversations about biliteracy, bilingual education and how to boost students' literacy outcomes.

This study warns against assuming that bilingual assessments are necessarily appropriate for high-stakes purposes. For instance, the lack of scalar invariance prevents using the scores from the Quarterly Assessments to establish meaningful comparisons between students from different linguistic groups. The policy recommendation is to **conduct extensive research on bilingual assessments to understand the conditions under which scalar invariance is reached**. Research on bilingual assessments can begin as an organic effort, by adding bilingual items onto large-scale national assessments and conducting regular psychometric analyses on them. Moreover, the invariance analysis shows that the sources of misfit are inconsistent across provincial groups, suggesting that bilingual assessments intended for high-stakes purposes could vary by region. The recommendation is to **incorporate some flexibility into the design of bilingual assessments to accommodate for structural differences between languages and students**. More research and data could clarify whether developmentally appropriate bilingual assessments can serve high-stakes purposes and meet the global monitoring demands, or should primarily serve formative and pedagogical functions. In the meantime, policy makers need to **expand the conversation around the meaning of literacy in linguistically diverse contexts such as the DRC and evaluate its implications for the global monitoring agenda**. More policy recommendations are shown in Table 13.

Table 13 Additional Recommendations Emerging from the Analysis of Bilingual Data

| Additional Recommendations |
|--|
| Conceive bilingual assessments as an organic element of bilingual programs. Bilingual assessments should mirror and support bilingual education programs. |
| Small-scale evaluation |
| Include examples of bilingual items in teaching materials, for example, at the end of each chapter. Train teachers on how to use these items to inform pedagogy, including remediation efforts. |
| Train other relevant school personnel - e.g., Head Teachers - on how to use bilingual assessments in formative ways. |
| Large-scale evaluation |
| Define linguistic subgroups of interest for large-scale monitoring and evaluation. |
| Publish detailed technical reports for each assessment used in large-scale monitoring and evaluation. |
| Conduct extensive item-level analyses on bilingual assessment data. Use the results of these analyses to inform key activities within bilingual education (trainings, curriculum revisions, etc.) |
| Continuously improve psychometric capacity and practice. For example, revise current assessment practices in relation to publicly available guidelines put forth by international organizations such as the International Test Commission (ITC). |

ACKNOWLEDGEMENTS

I would like to acknowledge the colleagues and institutions who supported this work. I must first thank the funders of the ACCELERE! 1 project, USAID and FCDO. Without their support this study would have not been possible. Similarly, I must thank the UNESCO Global Education Monitoring (GEM) Report Director, Manos Antoninis, and Team, for their decision to support research on bilingual assessments. I would also like to thank the senior management of ACCELERE! 1, particularly Sonia Arias, and the senior management at School-to-School International (STS), particularly Candace Debnam, for facilitating my involvement in this study. I want to acknowledge my STS colleagues who supported this work through their involvement in the ACCELERE! Quarterly Assessments: Hali Thomas, Selene Rangel, Stephanie Templeton, and Mark Lynd. Thank you! I also want to acknowledge Michel Rousseau for his important feedback and support. I am especially thankful to my UNESCO GEM Mentors, Daniel April and Lema Zekrya, for their multiple revisions of this article, their kindness, and availability. I also want to thank Priyadarshani Joshi, the GEM Team, and the rest of the 2021 Fellows for making this experience enjoyable and productive. Last, I want to thank my trilingual family, Nariman, Andes, and Josro, for teaching me more about bilingualism and its challenges than any textbook could have had.

REFERENCES

- Abedi, J. (2004). The No Child Left Behind act and English language learners: Assessment and accountability issues. *Educational Researcher*, 33(1), 4-14.
- Aber, J. L., Torrente, C., Starkey, L., Johnston, B., Seidman, E., Halpin, P., . . . Wolf, S. (2016). Impacts after one year of "Healing Classroom" on children's reading and math skills in the DRC: Results from a randomized control trial. *Journal of Educational Effectiveness*, 1-23.
- Adams, M. (1990). *Beginning to Read: Thinking and Learning About Print*. Cambridge: MIT Press.
- August, D., & Shanahan, T. (2006). *Developing literacy in second-language learners: Report of the National Literacy Panel on Language-Minority Children and Youth*. Mahwah, NJ : Lawrence Erlbaum Associates.
- Babaci-Wilhite, Z. (2013). Local languages of instruction as a right in education for sustainable development in Africa. *Sustainability*, 1994-2017.
- Babra Gora, R. (2017). (Re-)Integrating African languages into the Zimbabwean school curriculum. En E. Shizha, & N. Makuvaza, *Re-thinking postcolonial education in Sub-Saharan Africa in the 21st century* (págs. 141-158). Sense Publishers.
- Baker, C. (2011). *Foundations of bilingual education and bilingualism*. Bristol, UK: Multilingual Matters.
- Baker, C., & Prys Jones, S. (1998). Mother tongue. En C. Baker, & S. Prys Jones, *Encyclopedia of bilingualism and bilingual education* (págs. 47-50). Multilingual Matters.
- Becker, J., Rai, A., Ringle, C. M., & Völckner, F. (2013). Discovering unobserved heterogeneity in structural equation models to avert validity threats. *MIS Quarterly*, 37(3), 665-694.
- Benavot, A., & Tanner, E. (2008). *The growth of national learning assessments in the world, 1995-2006*. Background paper for the Education for all global monitoring report.
- Bokamba, E. (2011). Ukolonia in African Language Policies and Practices. *40th Annual Conference on African Linguistics* (págs. 146-167). Somerville, MA: Cascadilla Proceedings Project.
- Borg, P. (2015). *Rwanda's French-to-English "Language Switch": A discussion of Policy Motivations, Challenges and Implications*.
- Brown, T. (2015). *Confirmatory factor analysis for applied research*. The Guilford Press.
- Brown, T. A. (2015). CFA with Equality Constraints, Multiple Groups, and Mean Structures. En T. A. Brown, *Confirmatory Factor Analysis for Applied Research* (págs. 206-205). New York City: The Guilford Press.
- Buzasi, K. (2016). Languages and national identity in Sub-Saharan Africa: A multilevel approach. En M. Gazzola, & B. Wickström, *The Economics of Language Policy* (págs. 225–264). Cambridge, MA: MIT Press.
- Caesar, J. (2011). Language policy and the promotion of Kiswahili in Uganda. University of Dar es Salaam.

- Confemen. (2020). *PASEC 2019. QUALITÉ DES SYSTÈMES ÉDUCATIFS EN AFRIQUE SUBSAHARIENNE FRANCOPHONE*. Dakar, Senegal: Confemen.
- Cummins, J. (1979). *Cognitive/academic language proficiency, linguistic interdependence, the optimum age question and some other matters*. Toronto: OISE.
- Cummins, J. (1981). Empirical and theoretical underpinnings of bilingual education. *The Journal of Education*, 163 (1), 16-29.
- Cummins, J. (2001). Assessing second language proficiency among adults: Do we know what we are measuring? En J. Cummins, *Language, power and pedagogy: bilingual children in the crossfire* (págs. 86-111). Multilingual Matters Ltd.
- Cummins, J. (2001). Dilemmas of inclusions: integrating English language learners in standards-based reform. En J. Cummins, *Language, power and pedagogy: Bilingual children in the crossfire* (págs. 112-139). Multilingual Matters Ltd.
- Cummins, J. (2001). The threshold and interdependence hypotheses revisited. En J. Cummins, *Language, power and pedagogy: Bilingual children in the crossfire* (págs. 173-200). Multilingual Matters Ltd.
- Dekoke, T. (2016). Congolese migrants and South African language appropriation. *Language Matters*, 47 (1), 84-104.
- Depaepe, M., & Kikumbi, A. (2018). Educating girls in Congo: An unsolved pedagogical paradox since colonial times? *Policy Futures in Education*, 1-17.
- Depaepe, M., Simon, F., & Vinck, H. (2015). "New Education" for the Congo? The indigenist approach, new education and prescriptive pedagogy. *Prospects*, 45, 95-111.
- Dryden-Peterson, S., Adelman, E., Alvarado, S., Anderson, K., B. M., Brooks, R., . . . Suzuki, E. (2019). *Inclusion of refugees in national education systems*. Paris: UNESCO.
- Dubeck, M. M., Gove, A., & Alexander, K. (2016). School-based assessments: What and how to assess reading. En UNESCO-UIS, *Understanding what works in oral reading assessments* (págs. 42-57). Quebec, Canada: UNESCO-UIS.
- EDC. (2017). *Literacy, Language and Learning Initiative (L3) - Endline Report*. USAID.
- Edinga, K. (1999). La politique de l'éducation en république du Congo (Zaire) de 1960 a 1989.
- Ekkehard, W. (2018). 8. African socio- and applied linguistics: . En T. Güldemann, *The Languages and Linguistics of Africa*, (págs. 883-984). Berlin, Boston: De Gruyter Mouton.
- Eno, M., Ahad, A. M., & Shafat, A. (2019). Perceptions of academic administrators and policymakers on ESL/EFL education. *Journal of Somali Studies*, 6 (1), 93-119.

- Evans, R. (2018). Tshivenda-speaking teachers when Grade 4 learners transition to English. *The Journal for Transdisciplinary Research in Southern Africa*, 14 (2), 1-9.
- Gándara, F. (2017). Evaluating a Translingual Administration of the Early Grades Math Assessment (EGMA) in the Democratic Republic of the Congo. Doctoral Dissertations. 1086. .
- Gandara, F., & Randall, J. (2019). Assessing Mathematics Proficiency of Multilingual Students: The Case for Translanguaging in the Democratic Republic of the Congo. *Comparative Education Review*, 63 (1), 58-78.
- García, O. (2009). *Bilingual education in the 21st Century: A global perspective*. Malden, MA: Wiley-Blackwell.
- Gheondea-Eladi, A. (2015). Conducting critical literature reviews: A methodology note. *CALITATEA VIEII*, XXVI, 2, 167-190.
- Gibson, M. (2019). *Report on operations research for ACCELERE! 1: Sociolinguistic mapping and teacher language ability*. USAID.
- Global Partnership for Education. (October de 2021). *Democratic Republic of the Congo*. Obtenido de Where we work : <https://www.globalpartnership.org/where-we-work/democratic-republic-of-congo>
- Hedgcock, J., & Ferris, D. (2009). *Teaching readers of English*. New York, USA: Routledge.
- Heugh, K., & Mulumba, B. M. (2013). *An evaluation of the literacy and basic education (LABE) mother-tongue education project in UGANDA*.
- Hudson, R. A. (1996). *Sociolinguistics*. Cambridge: Cambridge University Press.
- Igboanusi, H. (2014). The English-only language education policy in The Gambia and low literacy rates. *International Journal of Bilingual Education and Bilingualism*, 7(5), 558-569.
- INS. (2019). *Enquête par grappes à indicateurs multiples, 2017-2018, rapport de résultats de l'enquête*. . Kinshasa: UNICEF.
- Jepkosgei Cheruiyot, C., & Kabellow, J. K. (2018). Challenges of enforcing English as the medium of instruction in upper primary classes in Kenya. *Journal of Education and Practice*, 9 (35), 24-32.
- Journal Officiel de la RDC. (2011). *Constitution de la République Démocratique du Congo*. Kinshasa: Cabinet du Président de la République.
- Kim, Y. G., & Piper, B. (2019). Component skills of reading and their structural relations: evidence from sub-Saharan African languages with transparent orthographies. *Journal of Research in Reading*, 42 (2), 326-348.
- Kolen, M., & Brennan, R. (2014). Linking. En M. Kolen, & R. Brennan, *Test Equating, Scaling and Linking* (págs. 487-536). Springer.
- Lazdowski, K. (2015). Stakeholders' roles in education language policy research in West Africa: A review of the literature. *Reconsidering Development*, 4 (1), 100-118.

- Louge, N., Gándara, F., Gibson, M., & Thomas, H. M. (In Press). *Teaching Linguistically Diverse Students: A Case Study for Developing Early Grade Literacy Materials in the Democratic Republic of the Congo*.
- Makhudu, P. (2017). A Study of African Postcolonial and Post-Apartheid Challenges: A Legacy of Ineptitude? *The 2nd Annual International Conference on Public Administration and Development Alternatives*, (págs. 376-388). Gaborone, Botswana.
- Makoni, S., & Makoni, B. (2015). Too many cooks spoil the broth. En W. E. Wright, S. Boun, & O. García, *The handbook of bilingual and multilingual education* (págs. 554-565). Wiley Blackwell.
- Makoni, S., & Mashiri, P. (2006). Critical historiography: Does language planning in Africa need a construct of language as part of its theoretical apparatus? . En S. Makoni, & A. Pennycook, *Disinventign and reconstituting languages* (págs. 62-89). Multilingual Matters.
- Management Systems International. (2019). *Policy linking method: Linking assessments to global standards*. Washington, D.C.: USAID. Obtenido de <http://gaml.uis.unesco.org/wp-content/uploads/sites/2/2019/08/GAML6-REF-1-Policy-linking-USAID.pdf>
- Marsh, H., Ludtke, O., Nagengast, B., Morin, A., & Von Davier, M. (2013, 18 (3)). Why item parcels are (almost) never appropriate: two wrongs do not make a right - camouflaging misspecification with item parcels in CFA models. *Psychological methods*, 257-284.
- Masagazi Masaazi, F., Ssentanda, M., & Ngaka, W. (2018). On Uganda's government's commitment to the development and implementation of the mother tongue education policy in post 2015-era. *Apples - Journal of Applied Language Studies*, 12 (2), 1-17.
- Meeuwis, M. (1999). Flemish nationalism in the Belgian Congo versus Zairian anti-imperialism: Continuity and discontinuity in language ideological debates. En J. Blommaert, *Language Ideological Debates* (págs. 381-424). Berlin: De Gruyter Mouton.
- Meeuwis, M. (2011). Bilingual inequality: Linguistic rights and disenfranchisement in late Belgian colonization. *Journal of Pragmatics*, 43, 1279-1287.
- Ministère de l'Enseignement Primaire Secondaire et Initiation à la Nouvelle Citoyenneté. (2015). *Stratégie sectorielle de l'éducation et de la formation 2016-2025*. Kinshasa: République Democratique du Congo.
- Ministere de l'Enseignement Primaire, Secondaire et Technique. (2021). Qualité de l'éducation : Première évaluation nationale des acquis scolaires par la CIEAS. Obtenido de <https://www.eduquepsp.education/v1/qualite-de-leducation-premiere-evaluation-nationale-des-acquis-scolaires-par-la-cieas/>
- Mokonzi, G., .Van Damme, J., De Fraine, B., Vitamara, P., Kimbuani, G., Mukiekie, A., . . . Bela, J. (2020). Educational Effectiveness Research in Africa: The Case of the Democratic Republic of the Congo (DRC). En *International*

Perspectives in Educational Effectiveness (págs. 185-207). Springer. Obtenido de https://doi.org/10.1007/978-3-030-44810-3_8

- Muaka, L. (2011). Language perceptions and identity among Kenyan speakers. *40th Annual Conference on African Linguistics* (págs. 217-230). Somerville, MA: Cascadilla Proceedings Project.
- Mwaniki, M., Arias, M., & Wiley, T. (2017). Bilingual education policy. En O. García, A. Lin, & S. May, *Bilingual and multilingual education, third edition* (págs. 35-50). Springer.
- Nishioka, S., & Durrani, N. (2018). Language and cultural reproduction in Malawi: Unpacking the relationship between linguistic capital and learning outcomes. *International Journal of Educational Research*, 1-12.
- Nthawakuderwa, B. (1985). Authenticity and problems of language planning in Zaire. *Language in education in Africa* (págs. 101-140). Edinburgh: University of Edinburgh.
- Nyarko Ansah, G. (2014). Re-examining the fluctuations in language in-education policies in post-independence Ghana. *Multilingual Education*, 4 (12), 1-15.
- Orwenjo, D. O. (2012). Multilingual education in Kenya: debunking the myths. *International Journal of Multilingualism*, 294-317.
- Páez, M. (2008). English language proficiency and bilingual verbal ability among Chinese, Dominican and Haitian immigrant students. *Equity and excellence in education*, 41 (3), 311-324.
- Penner Angrist, M. (2018). Language policy and girls' education in francophone West and Central Africa. *Social Politics*, 1-26.
- Piper, B., & Zuilkoski, S. (2015). Assessing reading fluency in Kenya: Oral or silent assessment? *International Review of Education*, 1-19.
- Pitsoe, V., & Machaisa, P. (2012). Teacher attrition catastrophe in Sub-Saharan Africa: A hurdle in the achievement of UPE, EFA policy goals and MDGs. *Science Journal of Sociology and Anthropology*, 1-7.
- Popham, J. (2001). Teaching to the test? *Helping All Students Achieve*, 58 (6), 16-20.
- Quin Yow, W., & Li, X. (2015). Balanced bilingualism and early age of second language acquisition as the underlying mechanisms of a bilingual executive control advantage: why variations in bilingual experiences matter. *Frontiers in Psychology*, 6, 1-12.
- Ramachandran, R. (2017). Language use in education and human capital formation: Evidence from the Ethiopian educational reform. *World Development*, 1-19.
- Rassool, N. (2014). Considering heteroglossia in language and development in Sub-Saharan Africa. *International Journal of the Sociology of Language*, 2014 (225), 51-73.

- Reddick, C., & Dryden-Peterson, S. (2021). Refugee education and medium of instruction: Tensions in theory, policy and practice. En C. Beson, & K. Leiden Konsonen, *Language Issues in Comparative Education* (págs. 208-233). The Netherlands: Brill | Sense.
- RTI International. (2015). *Early Grade Reading Assessment Toolkit*. Washington, D.C., USA: United States Agency for International Development.
- Samarin, W. (1987). *Lingua Franca*. Walter de Gruyter.
- Sanchez, S., Rodriguez, B., Soto-Huerta, M., Castro Villareal, F., Guerra, N., & Bustos Flores, B. (2013). A case for multidimensional bilingual assessment. *Language Assessment Quarterly*, 10 (2), 160-177.
- Sarstedt, M., Ringle, C. M., & Hair, J. F. (2017). Treating Unobserved Heterogeneity in PLS-SEM: A Multi-method Approach. En H. Latan, & R. Noonan, *Partial Least Squares Path Modeling* (págs. 197-217). Springer International Publishing.
- Schaefer, M., & Kotze, J. (2019). Early reading skills related to Grade 1 English second language literacy in rural South African schools. *South African Journal of Childhood Education*, 2223-7674.
- School-to-School International. (2019). *2018 Early Grade Reading Monitoring Assessment in the Democratic Republic of the Congo*. Washington, D.C.: USAID.
- Schroeder, L. (2012). Teaching and assessing independent reading skills in multilingual African countries: not as simple as abc. En C. Benson, & K. Kosonen, *Language Issues in Comparative Education*.
- Seid, Y. (2016). Does learning in mother tongue matter? Evidence from a natural experiment in Ethiopia. *Economics of Education Review* 55, 21-38.
- Shi, D., Lee, T., & Maydeu-Olivares, A. (2019). Understanding the Model Size Effect on SEM Fit Indices. *Educational and Psychological Measurement*, 310-334.
- Shohamy, E. (2006). Imagined multilingual schools: How come we don't deliver. En O. Garcia, T. Skutnabb-Kangas, & M. E. Torres-Guzman, *Imagining multilingual schools* (págs. 171-183). New York and Tronninge Mose: Multilingual Matters LTD.
- Shohamy, E. (2007). Language tests as language policy tools. *Assessment in Education Principles Policy and Practice*, 14(1), 117-130.
- Shohamy, E. (2011). Assessing multilingual competencies: Adopting construct valid assessment policies. *The Modern Language Journal*, 95 (iii), 418-429.
- Sibomana, E. (2020). Transitioning from a local language to English as a medium of instruction: Rwandan teachers' and classroom-based perspectives. *International Journal of Bilingual Education and Bilingualism*, 1-17.
- Smith, T., & McMillan, B. (2001). A Primer of Model Fit Indices in Structural Equation Modeling. *Annual Meeting of the Southwest Educational Research Association*. New Orleans.

- Technical Cooperation Group (TCG). (2018). *S.D.G. Indicator 4.5.2: Recommended Approaches for Measurement*. Mexico City, Mexico.
- The World Bank. (2015). *Democratic Republic of Congo: Education Sector Public Expenditure Review*. World Bank Group.
- The World Bank. (October de 2021). *Population - Total: Democratic Republic of the Congo*. Obtenido de The World Bank Data: <https://data.worldbank.org/indicator/SP.POP.TOTL?locations=CD>
- The World Bank. (October de 2021). *The World Bank in DRC*. Obtenido de Where we work: <https://www.worldbank.org/en/country/drc/overview#1>
- Trapido, J. (2015). Africa's leaky giant. *New Left Review*, 92 (II). Obtenido de <http://newleftreview.org/II/92/joe-trapido-africa-s-leaky-giant>
- Trudell, B. (2016). *The impact of language policy and practice on children's learning: Evidence from eastern and southern Africa*. UNICEF.
- UIS-UNESCO. (2017). *SDG 4 monitoring: framework development*. Madrid: UIS-UNESCO.
- UIS-UNESCO. (2018). *SDG 4 Data Digest: Data to Nurture Learning*. Montreal: UNESCO Institute for Statistics.
- UNESCO. (2017). *Accountability in education: meeting our commitments; Global education monitoring report, 2017/8*. UNESCO.
- UNESCO. (2020). *Global education monitoring report, 2020: Inclusion and education: all means all*. UNESCO.
- UNESCO. (2021). *Global education monitoring report, 2021/2: non-state actors in education: who chooses? who loses?* UNESCO.
- UNICEF. (2017). *Collecting data on foundational learning skills & parental involvement in education*. MICS UNICEF.
- United Nations. (2015). *Indicators and a Monitoring Framework for the Sustainable Development Goals: Launching a data revolution for the SDGs*. United Nations.
- United States Agency for International Development. (2019). *Education Reporting Guidance*. Washington, DC: USAID. Obtenido de https://www.edulinks.org/sites/default/files/media/file/Education_Reporting_Guidance_2020%20Updates_2020.10.16.pdf
- USAID. (2019). *Global Proficiency Framework: Reading and Mathematics Grades 2 to 6*. Obtenido de <https://gaml.uis.unesco.org/wp-content/uploads/sites/2/2019/05/GAML6-REF-16-GLOBAL-PROFICIENCY-FRAMEWORK.pdf>
- van de Vijver, F., & Leung, K. (1997). *Methods and data analysis for cross-cultural research*. London: SAGE Publications.

- van Gelderen, A., Schoonen, R., d. G., Hulstijn, J., Snellings, P., Simis, A., & Stevenson, M. (2003). Roles of linguistic knowledge, metacognitive knowledge and processing speed in L3, L2 and L1 reading comprehension: A structural equation modeling approach. *The International Journal of Bilingualism*, 7 (1), 7-25.
- Vandeyar, S., & Catalano, T. (2020). Language and identity: Multilingual immigrant learners in South Africa. *Language Matters*, 51 (2), 106-128.
- Vandeyar, S., & Vandeyar, T. (2011). Articulating cultures: Socio-cultural experiences of black female immigrant students in South African schools. *Gender & Behaviour*, 9 (2), 4161-4188.
- Velasco, P., & Garcia, O. (2014). Translanguaging and the writing of bilingual learners. *Bilingual research journal: The journal of the national association for bilingual education*, 37(1), 6-23.
- Williams, T. P. (2021). *Why Did Rwanda Abruptly Change the Language in Schools—Again?* Obtenido de <https://www.worldpoliticsreview.com/articles/29440/in-rwanda-language-change-in-schools-leaves-students-and-teachers-struggling>
- World Bank. (2018). *World Development Report 2018: Learning to Realize Education's Promise*. Washington, DC: World Bank. doi:doi:10.1596/978-1-4648-1096-1
- Zsiga, E., & Boyer, O. (2014). *Languages in Africa: Multilingualism, Language Policy, and Education*. Georgetown University Press.

APPENDICES

Appendix 1: SEM Results

Appendix Table 1 Model Comparison by Province

| | χ^2 | df | RMSEA | CFI | TLI |
|--------------------------|----------|----|-------|------|------|
| Ciluba | | | | | |
| Unidimensional | 181 | 31 | 0.11 | 0.95 | 0.92 |
| Bidimensional | 155 | 29 | 0.10 | 0.95 | 0.93 |
| Tridimensional | 154 | 28 | 0.10 | 0.95 | 0.93 |
| Bidimensional orthogonal | 366 | 31 | 0.16 | 0.88 | 0.82 |
| Lingala | | | | | |
| Unidimensional | 117 | 31 | 0.10 | 0.83 | 0.76 |
| Bidimensional | 105 | 29 | 0.10 | 0.85 | 0.77 |
| Tridimensional | 93 | 28 | 0.09 | 0.87 | 0.80 |
| Bidimensional orthogonal | 189 | 31 | 0.13 | 0.69 | 0.55 |
| Kiswahili | | | | | |
| Unidimensional | 303 | 31 | 0.10 | 0.88 | 0.82 |
| Bidimensional | 204 | 29 | 0.08 | 0.92 | 0.88 |
| Tridimensional | 204 | 28 | 0.08 | 0.92 | 0.87 |
| Bidimensional orthogonal | 325 | 31 | 0.10 | 0.87 | 0.81 |

Appendix Table 2 Fit Analysis Ciluba

| Ciluba | | | | | | | | |
|-------------------------------------|----------------|---------|---------------|---------|----------------|---------|------------|---------|
| | Unidimensional | | Bidimensional | | Tridimensional | | Orthogonal | |
| Parcel | Loading | P-Value | Loading | P-Value | Loading | P-Value | Loading | P-Value |
| Parcel 1 - Frequent French Words | 1.00 | | 1.00 | | 1.00 | | 1.00 | |
| Parcel 2 - Frequent French Words | 0.86 | 0.00 | 0.94 | 0.00 | 0.93 | 0.00 | 1.69 | 0.00 |
| Parcel 3 - ORF in Ciluba | 1.41 | 0.00 | 1.00 | | 1.00 | | 1.00 | |
| Parcel 4 - ORF in Ciluba | 1.76 | 0.00 | 1.25 | 0.00 | 1.25 | 0.00 | 1.24 | 0.00 |
| Parcel 5 - ORF in Ciluba | 1.74 | 0.00 | 1.23 | 0.00 | 1.23 | 0.00 | 1.15 | 0.00 |
| Parcel 6 - ORF in Ciluba | 1.88 | 0.00 | 1.33 | 0.00 | 1.33 | 0.00 | 1.24 | 0.00 |
| Parcel 7 - Writing words - FR | 0.75 | 0.00 | 1.00 | | 1.00 | | 1.21 | 0.00 |
| Parcel 7 - Writing words - NL | 0.69 | 0.00 | 0.95 | 0.00 | 0.95 | 0.00 | 0.51 | 0.00 |
| Parcel 8 - Grouping Words in French | 0.63 | 0.00 | 0.62 | 0.00 | 0.62 | 0.00 | 0.76 | 0.00 |
| Parcel 9 - Writing conventions | 0.45 | 0.00 | 0.47 | 0.00 | 0.45 | 0.00 | 0.32 | 0.00 |
| Percentage of Misfit/Low Fit | 0% | | 0% | | 0% | | 13% | |

Appendix Table 3 Fit Analysis Lingala

| Lingala | | | | | | | | |
|----------------------------------|----------------|---------|---------------|---------|----------------|---------|------------|---------|
| | Unidimensional | | Bidimensional | | Tridimensional | | Orthogonal | |
| Parcel | Loading | P-Value | Loading | P-Value | Loading | P-Value | Loading | P-Value |
| Parcel 1 - Frequent French Words | 1.00 | | 1.00 | | 1.00 | | 1.00 | |
| Parcel 2 - Frequent French Words | 1.06 | 0.00 | 0.93 | 0.00 | 0.93 | 0.00 | 0.80 | 0.00 |
| Parcel 3 - ORF in Ciluba | -0.71 | 0.00 | 1.00 | | 1.00 | | 1.00 | |
| Parcel 4 - ORF in Ciluba | -0.51 | 0.01 | 0.78 | 0.00 | 0.75 | 0.00 | 1.11 | 0.00 |
| Parcel 5 - ORF in Ciluba | 1.71 | 0.00 | -2.36 | 0.00 | -2.43 | 0.00 | -1.64 | 0.18 |

| | | | | | | | | |
|-------------------------------------|------------|------|------------|------|------------|------|------------|------|
| Parcel 6 - ORF in Ciluba | -0.64 | 0.00 | 1.08 | 0.00 | 1.01 | 0.01 | 2.89 | 0.17 |
| Parcel 7 - Writing words - FR | 0.38 | 0.00 | 1.00 | | 1.00 | | 0.17 | 0.12 |
| Parcel 7 - Writing words - NL | 0.56 | 0.00 | 0.68 | 0.00 | 0.57 | 0.00 | -0.26 | 0.01 |
| Parcel 8 - Grouping Words in French | 0.77 | 0.00 | 0.65 | 0.00 | 0.67 | 0.00 | 0.48 | 0.10 |
| Parcel 9 - Writing conventions | 0.01 | 0.98 | -0.02 | 0.95 | -0.24 | 0.32 | 0.05 | 0.76 |
| Percentage of Misfit/Low Fit | 56% | | 29% | | 29% | | 50% | |

Appendix Table 4 Fit Analysis Kiswahili

| Kiswahili | | | | | | | | |
|-------------------------------------|----------------|---------|---------------|---------|----------------|---------|------------|---------|
| Parcel | Unidimensional | | Bidimensional | | Tridimensional | | Orthogonal | |
| | Loading | P-Value | Loading | P-Value | Loading | P-Value | Loading | P-Value |
| Parcel 1 - Frequent French Words | 1.00 | | 1.00 | | 1.00 | | 1.00 | |
| Parcel 2 - Frequent French Words | 0.52 | 0.00 | 0.83 | 0.00 | 0.84 | 0.00 | 1.81 | 0.00 |
| Parcel 3 - ORF in Ciluba | 1.64 | 0.00 | 1.00 | | 1.00 | | 1.00 | |
| Parcel 4 - ORF in Ciluba | 2.14 | 0.00 | 1.35 | 0.00 | 1.35 | 0.00 | 1.34 | 0.00 |
| Parcel 5 - ORF in Ciluba | 1.92 | 0.00 | 1.22 | 0.00 | 1.22 | 0.00 | 1.08 | 0.00 |
| Parcel 6 - ORF in Ciluba | 1.49 | 0.00 | 0.95 | 0.00 | 0.95 | 0.00 | 0.85 | 0.00 |
| Parcel 7A - Writing words - FR | 0.89 | 0.00 | 1.00 | | 1.00 | | 0.48 | 0.00 |
| Parcel 7B - Writing words - NL | 0.28 | 0.00 | 1.63 | 0.00 | 1.64 | 0.00 | 1.02 | 0.00 |
| Parcel 8 - Grouping Words in French | 0.46 | 0.00 | 0.58 | 0.00 | 0.58 | 0.00 | 0.93 | 0.00 |
| Parcel 9 - Writing conventions | 0.67 | 0.00 | 1.33 | 0.00 | 1.33 | 0.00 | 0.37 | 0.00 |
| Percentage of Misfit/Low Fit | 11% | | 0% | | 0% | | 13% | |

Appendix Table 5 Multigroup Analysis, Linguistic Diversity at Home, Ciluba

| Ciluba | | | | | |
|-------------------------------|----------|----|-----------------|-------------|---------|
| | χ^2 | df | χ^2_{diff} | Δ df | P-Value |
| Measurement invariance | | | | | |
| Configural | 202 | 58 | | | |
| Metric | 236 | 67 | 34 | 9 | 0.00 |
| Scalar | 334 | 74 | 99 | 7 | 0.00 |

Appendix Table 6 Multigroup Analysis, Linguistic Diversity at Home, Lingala

| Lingala | | | | | |
|-------------------------------|----------|----|-----------------|-------------|---------|
| | χ^2 | df | χ^2_{diff} | Δ df | P-Value |
| Measurement invariance | | | | | |
| Configural | 141 | 58 | | | |
| Metric | 175 | 67 | 34 | 9 | 0.00 |
| Scalar | 218 | 74 | 43 | 7 | 0.00 |

Appendix Table 7 Multigroup Analysis, Linguistic Diversity at Home, Kiswahili

| Kiswahili | | | | | |
|-------------------------------|----------|----|-----------------|-------------|---------|
| | χ^2 | df | χ^2_{diff} | Δ df | P-Value |
| Measurement invariance | | | | | |
| Configural | 245 | 58 | | | |
| Metric | 269 | 67 | 24 | 9 | 0.00 |

| | | | | | |
|--------|-----|----|----|---|------|
| Scalar | 305 | 74 | 35 | 7 | 0.00 |
|--------|-----|----|----|---|------|

Appendix Table 8 Multigroup Analysis, Type of Language, Ciluba

| Ciluba | χ^2 | df | χ^2_{diff} | Δ df | P-Value |
|-------------------------------|----------|----|-----------------|-------------|---------|
| Measurement invariance | | | | | |
| Configural | 183 | 56 | | | |
| Metric | 213 | 65 | 30 | 9 | 0.00 |
| Scalar | 224 | 71 | 11 | 6 | 0.10 |

Appendix Table 9 Multigroup Analysis, Type of Language, Lingala

| Lingala | χ^2 | df | χ^2_{diff} | Δ df | P-Value |
|-------------------------------|----------|----|-----------------|-------------|---------|
| Measurement invariance | | | | | |
| Configural | 135 | 56 | | | |
| Metric | 139 | 65 | 4 | 9 | 0.90 |
| Scalar | 172 | 71 | 33 | 6 | 0.00 |

Appendix Table 10 Multigroup Analysis, Type of Language, Kiswahili

| Kiswahili | χ^2 | df | χ^2_{diff} | Δ df | P-Value |
|-------------------------------|----------|----|-----------------|-------------|---------|
| Measurement invariance | | | | | |
| Configural | 247 | 56 | | | |
| Metric | 258 | 65 | 11 | 9 | 0.30 |
| Scalar | 297 | 71 | 39 | 6 | 0.00 |

Appendix Table 11 Invariance Analysis, Fit on Focal Group, Linguistic Diversity

| | χ^2 | df | RMSEA | CFI | TLI |
|------------------|----------|----|-------|------|------|
| Ciluba | | | | | |
| One language | 120 | 29 | 0.13 | 0.94 | 0.90 |
| Lingala | | | | | |
| One language | 53 | 28 | 0.12 | 0.83 | 0.73 |
| Kiswahili | | | | | |
| One language | 49 | 29 | 0.06 | 0.96 | 0.94 |

Appendix Table 12 Invariance Analysis, Fit on Focal Group, Minority Languages

| | χ^2 | df | RMSEA | CFI | TLI |
|--------------------|----------|----|-------|------|------|
| Ciluba | | | | | |
| Minority languages | 38 | 29 | 0.09 | 0.96 | 0.94 |
| Lingala | | | | | |
| Minority languages | 34 | 28 | 0.07 | 0.98 | 0.96 |
| Kiswahili | | | | | |
| Minority languages | 120 | 29 | 0.09 | 0.89 | 0.82 |