

University of Massachusetts Amherst
ScholarWorks@UMass Amherst

Doctoral Dissertations

Dissertations and Theses

November 2017

Evaluating a Translingual Administration of the Early Grades Math Assessment (EGMA) in the Democratic Republic of the Congo

Fernanda Gandara

Follow this and additional works at: https://scholarworks.umass.edu/dissertations_2



Part of the [Anthropological Linguistics and Sociolinguistics Commons](#), [Bilingual, Multilingual, and Multicultural Education Commons](#), [Educational Assessment, Evaluation, and Research Commons](#), [Educational Methods Commons](#), [Other Linguistics Commons](#), and the [Science and Mathematics Education Commons](#)

Recommended Citation

Gandara, Fernanda, "Evaluating a Translingual Administration of the Early Grades Math Assessment (EGMA) in the Democratic Republic of the Congo" (2017). *Doctoral Dissertations*. 1086.
https://scholarworks.umass.edu/dissertations_2/1086

This Open Access Dissertation is brought to you for free and open access by the Dissertations and Theses at ScholarWorks@UMass Amherst. It has been accepted for inclusion in Doctoral Dissertations by an authorized administrator of ScholarWorks@UMass Amherst. For more information, please contact scholarworks@library.umass.edu.

Evaluating a Translingual Administration of the Early Grades Math Assessment (EGMA)
in the Democratic Republic of the Congo

A Dissertation Presented

by

FERNANDA GÁNDARA

Submitted to the Graduate School of the
University of Massachusetts Amherst in partial fulfillment
of the requirements for the degree of

DOCTOR OF PHILOSOPHY

September 2017

College of Education

Research, Educational Measurement and Psychometrics

© Copyright by Fernanda Gándara 2017
All Rights Reserved

Evaluating a Translingual Administration of the Early Grades Math Assessment (EGMA)
in the Democratic Republic of the Congo

A Dissertation Presented

By

FERNANDA GÁNDARA

Approved as to style and content by:

Jennifer Randall, Chair

Stephen G. Sireci, Member

Jonathan D. Rosa, Member

Joseph B. Berger, Senior Associate Dean
College of Education

DEDICATION

(Encountered in Drown by Junot Díaz)

The fact that I
am writing to you
in English
already falsifies what I
wanted to tell you.
My subject
How to explain to you that I
don't belong to English
though I belong nowhere else

Gustavo Pérez Firmat

ACKNOWLEDGEMENTS

Quisiera poder agradecer a todas las personas que de una forma u otra contribuyeron a este proceso, pero será muy difícil hacerlo en un par de planas. Quisiera primero agradecer a todos en la escuela de psicometría de UMass, su faculty (Ron, Jennifer, Lisa, Craig, Steve, April), sus estudiantes y visiting scholars (Louise, Alejandra, Joshua, Francis, Molly, Fen Fan, Xi, Vahab, Duy, Joseph, Minjeong, Denise, Javier, Nacho, Guler, Cristina, Gabriel), y su planta administrativa (Peg!), por haberme permitido disfrutar estos años de aprendizaje y de colaboración. Haber elegido UMass fue una bendición en muchísimos sentidos y me siento extremadamente satisfecha con el programa.

En particular, quisiera agradecer a mi gran comité. A Jennifer Randall, mi advisor, quien con disciplina y paciencia me acompañó de principio a fin: thanks a lot, ¡Jennifer! Quisiera agradecer a Steve Sireci, a quien considero mi amigo, un mentor académico, y quien me ayudó cuando más lo necesité: ¡muchas gracias querido Steve! Nunca olvidaré tu apoyo. Quisiera agradecer al tercer miembro del comité, Jonathan Rosa, por acceder a participar de este trabajo y por su brillante feedback: gracias Jonathan, sin ti este trabajo no sería del mismo nivel. Tengo un lujo de comité y no puedo estar más agradecida por ello.

La pasada por Amherst-Northampton no hubiera sido lo mismo sin mis queridos roommates (Javier, M'Liss, Zoe, Caroline, Josh, Manu, Nara) quienes me hicieron sentir a gusto y gracias a quienes puedo llamar Western MA mi segundo hogar. En los buenos y malos momentos, mis roommates me llenaron de buenas energías: ¡muchas gracias! Por otra parte, Western Mass no hubiera sido lo mismo sin los buenos amigos: al parche Chile - Colombia – España – Middle East – Econ – ALACS - REMP le debo grandes momentos. Mención especial a María Errea y Javier Suarez con quienes nos reímos a carcajadas, y a

Ana Ospina, amiga, vecina y confidente, quien me acompañó en los highs and lows. Mención doblemente especial para mi Nariman, quien con su inigualable chispa y sentido del humor me llena la vida de alegría, y quien con cariño y paciencia, me ayudó enormemente a concluir este proceso. ¡Grandes años en el Happy Valley!

Quiero agradecer a todos quienes me apoyaron desde la distancia: a mis papás, por enseñarme a trabajar muy duro, siempre. A la Mónica Silva, antigua jefa y amiga, por empujarme a estudiar este doctorado y por influenciar mi carácter – tenemos claras diferencias de pensamiento, pero nunca he conocido a alguien más firme e integra que la Mónica; a mis amigas de colegio y mis amigos de la vida, por estar siempre conmigo y por ser mi soporte emocional; al resto de mi familia, a mis antiguos colegas y jefes, y mis antiguos profesores, a todos quienes marcaron mi vida personal y profesional: muchas gracias por el apoyo y por acompañarme en los años previos a la realización de este doctorado. Quiero también agradecer a School-to-School International, a mis jefes y colegas, por darme todas las facilidades para terminar de escribir esta disertación.

Finalmente, quiero agradecer a todas las personas que me ayudaron con la recolección de datos para esta tesis: Julie, Willy, al equipo IRC-Congo, a las enumeradoras Natacha y Therese, a los choferes, a los médicos que me trataron la malaria y extirparon quistes, a los muchachos del hotel en el Río Congo, quienes hicieron que el tiempo volara, y por sobre todo, a las niñas y los profesores con quienes tuve el placer de trabajar.

Simplemente, gracias.

ABSTRACT

**EVALUATING A TRANSLINGUAL ADMINISTRATION OF THE EARLY
GRADES MATH ASSESSMENT (EGMA) IN THE DEMOCRATIC REPUBLIC
OF THE CONGO**

SEPTEMBER 2017

FERNANDA GÁNDARA,

INGENIERÍA CIVIL, PONTIFICIA UNIVERSIDAD CATÓLICA DE CHILE

Ph.D. UNIVERSITY OF MASSACHUSETTS AMHERST

Directed by: Professor Jennifer Randall

Translanguaging is a view around languages that normalizes diglossia without separation: the linguistic resources of the bilinguals are considered one integrated system. Translanguaging is also a language practice of bilinguals, who select features from their entire linguistic repertoire to make sense of the world around them. Translanguaging is widely used by students and teachers in the bilingual classroom, as it allows students to build upon their entire set of resources, enhance learning outcomes, perform identities, and develop their languages even further. However, translanguaging is rarely used in assessments of bilinguals. Assessments of bilinguals, especially large-scale tests, are typically monolingual in focus and not appropriate for a large portion of the population, who cannot perform as one or two monolinguals. While psychometricians and test developers have spent large amount of resources in developing and testing linguistic accommodations, their efforts are not entirely solving the problems faced by bilinguals.

Translanguaging is a framework that may overcome the limitations of linguistic accommodations. However, there is few research on how to properly implement translanguaging in assessments, particularly, in content assessments. The purpose of this work was to evaluate the effectiveness and appropriateness of implementing a translingual administration of the EGMA assessment in the region of Mbandaka, in the Democratic Republic of the Congo. Using a mixed-methods design, I looked at the effect of translanguaging on scores and the alignment of the framework with the classroom practices enounced by teachers. The results of this study show that the translingual version of the EGMA had a positive effect on the scores of girls who identified as bilinguals, and improved the reliability estimates of all the tasks. The results also show that the translingual EGMA is more appropriate for the context of Mbandaka, yet there are characteristics that prevent us from considering the test fully appropriate for the region. Further research must shed light on the particular aspects of the translingual administration that explained the improvements observed in this study. Future studies should also clarify potential routes to a better and more effective implementation of translanguaging in content assessments.

TABLE OF CONTENTS

ACKNOWLEDGEMENTS v

ABSTRACTvii

LIST OF TABLES xii

LIST OF FIGURES xiv

CHAPTER

1. INTRODUCTION 1

 1.1 Language Choices in Educational Settings 1

 1.2 Language as an Invention 5

 1.3 Language Issues with Educational Assessments 8

 1.4 Assessing the Content Knowledge of Emerging Bilinguals 12

 1.4.1 A Traditional Solution: Linguistic Accommodations 13

 1.4.2 An Alternative Framework to Imagine Assessments 16

 1.5 Statement of Purpose 18

2. LITERATURE REVIEW 23

 2.1 The Democratic Republic of the Congo 23

 2.2.1 Linguistic Debates in the Democratic Republic of the Congo 24

 2.2.2 Educational System in the Democratic Republic of the Congo... 28

 2.2.3 The Structure of the Educational System 31

 2.1.3.1 The Characteristics of Primary School 32

 2.1.3.2 Mathematics Instruction in Primary School 32

 2.2 Assessing Mathematics 35

 2.2.1 Academic Language as a Framework to Support Assessment
 Development..... 35

 2.2.2 The Academic Language of Mathematics 37

 2.3 Research on Linguistic Accommodations Relevant to Translanguaging in
 Assessments 39

3. METHODOLOGY	50
3.1 Overview	50
3.2 Research Design	52
3.2.1 The Early Grades Mathematics Assessment	52
3.2.1.1 Characteristics of the Core EGMA	53
3.2.2 A Translingual Administration of the EGMA Assessment.....	56
3.2.3 Sample	59
3.2.3.1 Sampling Methodology	59
3.2.4 Procedures	61
3.3 Data Analysis	64
3.3.1 First Research Question	64
3.3.2 Second Research Question	70
3.3.3 Third Research Question	71
4 RESULTS	74
4.1 Does a translingual administration of the EGMA assessment have an impact on math achievement of Congolese girls, as evidenced by a significant difference in their scores when compared to the traditional administration?	74
4.2 Does a translingual administration of the EGMA assessment allow us to obtain more reliable information about Congolese girls’ math knowledge, as evidenced by higher reliability estimates when compared to the those obtained from the traditional administration?	82
4.3 Is the translingual administration EGMA assessment appropriate for the context in which this data was collected?	85
4.3.1 The Context	85
4.3.2 Flexible Bilingualism Among the Assessed Girls	89
4.3.3 Feedback from Teachers	94
4.3.3.1 Best Evaluation Method	96
4.3.3.2 Demonstration of Understanding	96
4.3.3.3 Use of Evaluations	97
4.3.3.4 Language of Instruction	98
4.3.4 Feedback from Enumerators	100

5 DISCUSSION	102
5.1 Further Directions for Research	115
5.2 Limitations	122
APPENDICES	125
A. EGMA BASELINE FORM	125
B. EGMA MIDLINE FORM	128
REFERENCES	131

LIST OF TABLES

Table	Page
Table 1. Linguistic Accommodations by Type	14
Table 2. Intermediate Objectives for Mathematics Instruction in Primary Schooling	34
Table 3. Characteristic of the Selected Meta-Analyses	45
Table 4. Details of the Core EGMA	54
Table 5. Grade Level Distribution of the Girls	61
Table 6. Original Set of Questions for Teachers	63
Table 7. Original Set of Questions for Enumerators	64
Table 8. Coding Schemes Used in the Regression Analysis	70
Table 9. Summary of Methods by Research Question	73
Table 10. Test of Normality for Tasks and Total Scores by Type of Administration	77
Table 11. Kolmogorov-Smirnov and Anderson-Darling Tests by Type of Administration	78
Table 12. Results of the 2-Way ANOVA	79
Table 13. Results of the 3-Way ANOVA	79
Table 14. Simple Comparisons 3-Way ANOVA	80
Table 15. Unique Contribution of the TL and the HL Terms on the Regression	

Analysis	81
Table 16. Results for the Final Multiple Regression Model	81
Table 17. Adjusted Means	82
Table 18. Parallel Forms Reliability Estimates	83
Table 19. Proportion of Shared Variance between Forms	83
Table 20. Z-test for Differences in Correlations	84
Table 21. Lexical Items across Forms by Tasks	92
Table 22. Lexical Items by Language and Task Across Administrations	93

LIST OF FIGURES

Figure	Page
Figure 1. Age Distribution of Girls	61
Figure 2. Administration of the EGMA and EGMA TL	62
Figure 3. Differences in Scores between Forms by Administration Type	76
Figure 4. Distribution of Scores by Tasks and Totals	76
Figure 5. Adjusted Means – Graph	82
Figure 6. Number of Languages Used in Providing Answers	91
Figure 7. English Translation of the Interview to Enumerator	101

CHAPTER 1

INTRODUCTION

1.1 Language Choices in Educational Settings

Multilingualism – or the use of multiple languages by an individual or a community - is the norm and not the exception. And multilingualism creates challenges for social policy and institutions; in particular, multilingualism entails a need for language planning. Language planning refers to the intention to affect structure, function, and acquisition of languages in a given space (Tollefson, 2008). Modern societies require language planning to serve their linguistically diverse populations. The experience of linguistic majorities and minorities will be different if governments embrace linguistic pluralism – which promotes and extends linguistic diversity – than if they impose policies for assimilation of the majority language – which forces linguistic minorities to change their language practices (Baker, 2011). Societies need to address the increasing tension that exists between the attempt to homogenize certain linguistic practices and the increasing linguistic diversity of their populations (García, Skutnabb-Kangas, & Torres-Guzman, 2006). On top, societies need to plan language structure, function, and acquisition to meet modern socio-economic challenges. Globalization has changed the linguistic landscape beyond the natural consequences of increased migration trends. As an example, societies now face a rapid and invasive expansion of English as well as a commodification of language, which is now less of a sign of national identity and more a skill for a dynamic job market (Block, 2008; Tollefson, 2008). Speaking “languages”, especially English in a standard variety, is valued as a skill that global employees need to possess. Given that English is not the first language of most countries, governments need to make language decisions so as to enable

their citizens to participate in the globalized world, while paying attention to the issues of citizenship and diversity. On the contrary, in countries where English is the official language, it is used as a way to reproduce and reinforce nationalist discourses. The ways in which Governments solve their language challenges vary considerably. However, all states depend largely on social institutions to enact them.

In particular, educational institutions play a key role in language planning. There are a number of language decisions that affect schools. A clear example is that schools have to decide what language(s) of instruction to use. In the case of choosing multiple languages for instruction, there needs to be a definition with regards to the relationship that languages will have throughout school grades, subjects, and so forth. As Mackey noted in 1972, schools that use more than one language for instruction may vary with regards to: (a) the medium of instruction (single or dual medium), (b) the pattern and direction of the development of the languages, (c) the distribution of the languages, and (d) the pace and extent of the shift from one medium to another (Mackey, 1972). The decisions that schools take in this sense are affected by the ideologies under which they operate; schools work towards different language, cultural and social integration aims (Baker, 2011; García, 2009). Accordingly, scholars distinguish *multilingual schools* as those institutions that use more than one language of instruction and that embrace a broader goal of educating equitably and to build on the linguistic and cultural diversity of the learners (García, 2009; García et al., 2006). The language policy of multilingual schools is such that diverse languages are valued and used as a resource not as a barrier; in these schools, biliteracy is often an outcome (García, 2009). Multilingual schools operate under different assumptions than *immersion schools* – or schools where language minorities are forced into the

mainstream language (García, 2009; García et al., 2006). Immersion schools do not value linguistic or cultural diversity, and therefore, neglect the practices of sub-groups that deviate from the mainstream. The experiences and outcomes of multilingual and immersion schools should be very different; schools and educational institutions have the power to shape the linguistic practices that a population will adopt.

Literature offers some frameworks to classify school programs according to their linguistic characteristics. In the case of bilingual education, we may classify programs into three categories: *monolingual forms of bilingual education*, *weak forms of bilingual education*, and *strong forms of bilingual education* for bilingualism and literacy (Baker, 2011). Monolingual forms of bilingual education include programs where the aim is to foster monolingualism and where there is one language valued and promoted, often that of the mainstream. Under these programs, there is no room for second language practices, and languages other than that valued by the school are excluded from the educational realm. Immersion schools would be classified into this category. On the other hand, weak forms of bilingual education encompass programs that encourage bilingual practices, such as learning a foreign language, but where the outcome is rarely bilingualism or biliteracy. An example of such programs are the *transitional models of bilingual education* – or those models where language minority students are allowed to use their language in instruction for a limited number of years. The languages of the minorities are not entirely neglected, yet the linguistic goal is to produce a language shift towards that of the mainstream. In turn, *strong forms of bilingual education* encompass programs where the aim is that students become bilingual, regardless if they serve language minorities or majorities (Baker, 2011).

These programs encourage and value cultural and linguistic pluralism. Schools adopting strong forms of bilingual education are those technically considered multilingual.

From an instructional perspective, the choice of bilingual model of education would need to be selected to guarantee an optimal preparation of students for life and citizenship. However, language policies in education do not necessarily match best instructional practices. For example, research shows that strong forms of bilingual education may be preferred in terms of educational outcomes (Baker, 2011; García, 2009). Such programs enable students to become bilingual, an outcome that is often accompanied by an increase in achievement across the curriculum (Baker, 2011; García, 2009). In addition, non-native students take at least five years to catch up with their native peers in terms of language proficiency, suggesting that monolingual models of education or some weak forms of bilingual education, may be detrimental for this population (Baker, 2011; Shohamy, 2006, 2011). Thus, a pedagogic lens favors strong forms of bilingual education, as they benefit students' achievement and better prepare them to face the contemporary reality. However, strong forms of bilingual education are rare in contemporary schooling (Baker, 2011; Shohamy, 2006). The scarcity of strong bilingual education programs is partly due because of ideological reasons, as language is seen as a powerful and effective device to strengthen national identity (Shohamy, 2006; Tollefson, 2008). Politicians and decision makers may be more concerned about issues of citizenship and identity than of learning outcomes. In educational language matters, such as the choice of bilingual education model, politics often proves stronger than research (Baker, 2011; García et al., 2006; Shohamy, 2006).

Language planning in education is not solely a pedagogical matter. Language policies do not operate in isolation and are inseparable from the socio-economic and

political framework in which educational institutions are situated. Historically, language policies have served nationalist political agendas by imposing a standard variety of a national language as the “correct language” (Garcia et al., 2006; Shohamy, 2006). Institutions, including educational, formally or informally impose the grammar, lexicon and pronunciation of the native speakers as the legitimate language to be used (García et al., 2006; Shohamy, 2006; Tollefson, 2008). These practices exclude or de-legitimize language deviations from such norms, leading to the marginalization of language minorities and the creation or maintenance of social inequalities (García et al., 2006; Shohamy, 2006). The exclusion of certain groups from social practices has obvious consequences for their self-esteem and well-being. In particular, the language practices of schools contain messages and establish hierarchies that impact students’ identities (García, 2009; García et al., 2006; Shohamy, 2006; Tollefson, 2008). But despite the fact that monolingual practices in education may be negative at the social and individual level, they are still a common practice (García et al., 2006); strengthening national identity may be more relevant than learning outcomes or the social integration of language minorities. Language planning in schools is both an ideological and pedagogical matter, and it needs to be analyzed with considerations to the broader context in which it operates (García et al., 2006).

1.2 Language as an Invention

There is some consensus that language in education is not solely an educational matter, since language itself is a political tool. Languages have been instrumental to various socio-political projects at different points in time. For example, language served the creation of nation-states, by providing a cultural basis upon which to imagine the

corresponding political community (Anderson, 1983). Language also served colonialist endeavors by rendering itself a tool of tangible political expansion. Missionaries and colonial officers used language to accomplish their evangelizing and administrative duties (García, 2009). It is no coincidence that people from former French colonies speak at least some form of French, or that English is one of the official languages in most former British colonies. The power of languages as a form of control and hegemony is undeniable.

However, acknowledging that language is power is not enough to overcome it. In most debates or reflections about the topic, language is typically treated as an autonomous system, as something that exists outside and above human beings. Makoni and Pennycook (2007) bring light to the flawed nature of this conceptualization. The authors challenge the notion of languages as autonomous systems by stating that some schools of thought (e.g. integrational linguistics) do not require postulating such thing as part of their linguistic theory. They also say that any hard science approach to linguistics would focus on how people communicate rather than on the idea of languages. The concept of language itself relies on a series of unjustified special assumptions, which prevents it from becoming the essential object of (hard) scientific analysis. Moreover, the authors prove that the statement of “languages exist as autonomous systems” does not make sense under approaches that focus on language ideologies. If comprehending language requires acknowledging the beliefs of the participants, we simply cannot state that language is independent of its context. Makoni and Pennycook (2007) argue that not only languages are not autonomous systems, but that they are actually inventions. To become aware of the invented nature of languages, and to understand the meta-discursive, ideological, and historical origins of such inventions, are to them, the essential steps to overcome our current limitations in

linguistic debates. The authors suggest that a further step in the attempt to uncover and undo the political power of languages is to come to good terms with the idea that language is an invention.

The idea of language as an invention makes sense. For example, the fact that there is no consensus on the number of languages is clear evidence that languages are socially constructed (García, 2009; Makoni & Mashiri, 2007). The invention of languages is particularly evident when studying colonialism. Colonialism refers to the project of European political domination that took place between the 16th and 20th century, after the national liberation movements in 1960 (Kohn, 2014). Language was an instrument that missionaries and colonial administrators used to carry out their agendas. In order to effectively fulfill their purposes, missionaries and colonial linguistics embarked in a project of creating languages: constructing grammars, orthographies, dictionaries, even language academies (García, 2009). Language as an invention does not mean that people didn't have language before colonialism, but that the notion of language and the characteristics of the "languages" that were used to describe, define, and control colonized populations, were created after European, Western, Colonial, and Christian ideologies and meta-discursive regimes. Missionaries and colonialists created languages that resembled their own views and reproduced their own semiotic systems.

The idea of language as an invention is relevant to postcolonial and contemporary contexts (Makoni & Pennycook, 2007; Makoni & Mashiri, 2007). The colonial and/or conventional Western linguistics ideologies that the "ideology of invention" denounces, are still present in our conceptualizations of language. And we cannot avoid this problem because despite being invented, the effects of language are very real (Makoni &

Pennycook, 2007). Conventional ideas about languages or about language continue to oppress certain populations. For example, Branson and Miller (2007) explain how using a conventional approach to analyze a language known as Kata Kolok, a sign language used in Bali, may be detrimental. The Kata Kolok does not meet the Western and academic criteria of what a language should look like. For example, the signs used in the Kata Kolok are not arbitrary (Branson & Miller, 2007), so the Kata Kolok does not meet the criteria of autonomy that traditional linguistics preaches. The concrete effect of this discrepancy is that any traditional analysis of the Kata Kolok is likely to result in some form of epistemic violence, in some form of linguistic oppression. And calling for more linguistic diversity or respect for linguistic rights would not solve the matter. It is for these reasons that researchers like Makoni and Pennycook call for a *disinvention* of our notions about language. Understanding that languages are socially and politically constructed is essential to face situations in which there are reasons to change them or the way we think about them. Consequently, and concretely, the ideology of invention creates an additional layer of complexity to any analysis of language in education.

1.3 Language Issues with Educational Assessments

An aspect of language planning that deserves special attention is the language of educational tests. Tests are devices or procedures which sample behaviors in a specified domain and these behaviors are evaluated accordingly (American Educational Research Association, American Psychological Association, & National Council for Measurement in Education, 2014). There are many types of tests, but typically, they consist of a set of questions to which examinees need to provide answers in either a written or oral response. Regardless of their nature and purpose, all tests are subject to common decisions with

regards to language. Some language decisions affecting tests include (a) the choice of code of the questions or procedures being used to collect behaviors, (b) the choice of code in which responses may be provided, and (c) the norms used in the evaluation or scoring procedures. These language choices are means to shape behaviors and set expectations that cannot be overlooked. Testing may further enforce educational language policies (Shohamy, 2011, 2006) and so the language of tests should be carefully evaluated.

One of the strong criticisms to the language of educational tests emerges from the fact that most of them – at least in the context of standardized large-scale assessments – do not recognize the wide linguistic variety of test takers. Standardized assessments are typically *monolingual* in their focus (Escamilla, 2006; García, 2009; Gottlieb, 2014; Turkan, & Guzman-Orth, 2017; Shohamy, 2006; 2009). In other words, these assessments are developed as if the language of test takers was the same, as if they all used homogeneous and standard linguistic practices. The monolingual approach to test construction is reflected in the language of the questions, the language requirements for the answers, and the language norms of the scoring rubrics. Such approach to test development does not properly address the reality of *emerging bilinguals* – or students who through schooling and through acquiring the native language of the country they live in, become bilingual, being able to operate in their home language and in the new language (García, Kleifgen, & Falchi, 2008). In the classrooms, emerging bilinguals engage in complex and flexible linguistic practices, using resources from more than “one language” (e.g. Creese & Blackledge, 2010). Monolingual tests do not account for these practices, and therefore, do not *align* to the instruction of emerging bilinguals. *Alignment* refers to the degree to which different components of an educational system work in conjunction towards the

achievement of a particular goal (Martone & Sireci, 2009). While the language component is largely absent in alignment methodologies, it should be an element to consider when evaluating the extent to which tests and instruction are mutually supportive. As Abedi (2004) emphasizes, the language of the tests needs to match the language of instruction. In a disinvention effort, we should reinterpret this statement as “the linguistic practices allowed and encouraged in assessments should match the linguistic practices that students deploy in school settings”. Monolingual tests do not meet this requirement; indeed, they contradict the reality of the bilingual classrooms (García, 2009) and the current beliefs around teaching multilingual students held by many in the field of language education (Shohamy, 2011).

Monolingual tests are also questioned on ideological grounds. Monolingual assessments relate to a *language as a problem* viewpoint, which favors using one majority language as means to increase integration, cohesiveness, and to diminish socio-political or even economic problems (Baker, 2011). These assessments enforce policies that promote the use of one language, typically that of the majority. Most remarkably, monolingual assessments used in bilingual contexts fail the most fundamental measurement endeavor: these assessments cannot appropriately measure the proficiency of emerging bilinguals. In the case of language proficiency assessments, it is unlikely that emerging bilinguals can do as well as native speakers because they are still learning the language of the assessment (García, 2009). In the case of content assessments, monolingual approaches make it hard to figure out to what extent the performance of these students is a matter of content knowledge or a matter of language proficiency (Abedi, 2004; García, 2009; Shohamy, 2011). These technical shortcomings raise fairness concerns. Researchers have

documented a number of detrimental effects of these assessments on emerging bilingual students, including disproportionate rates of emerging bilinguals being placed into remedial education, education programs, or low curriculum tracks, at least in the U.S. (García, 2009, García et al., 2008; Valenzuela, 2005; Yzquierdo, 1995). Assessing bilinguals as monolinguals leads to inappropriate conclusions regarding their knowledge and skills. And in the case of high-stakes assessments, misdiagnoses may lead to irreversible negative consequences for this population. Monolingual assessments of bilinguals are a form of epistemic violence and cultural and linguistic oppression.

Multilingual schools and educational systems in general, need better assessment practices for emerging bilingual students. Bottom line, assessments should allow children to demonstrate what they know and are learning in all of their languages (Escamilla, 2006; García, 2009). Some have espoused that it is conceivable to offer assessment opportunities in different languages. The argument is that emerging bilinguals perform better when assessed in their home languages (e.g. Escamilla, 2006). Notwithstanding, these practices may not be completely appropriate; bilingual students do not perform as two monolinguals (Shohamy, 2011); the language proficiency of a bilingual student is not comparable to the proficiency of monolinguals in each of the corresponding languages. In addition, the assumption that emerging bilinguals have high proficiency in their home languages – or even that they may achieve such proficiency - is one that cannot always be met (García, 2009; García et al., 2008; Shohamy, 2011). Moreover, using monolingual assessments in different languages brings additional difficulties, as valid test translations are hard to produce (Logan Terry & Wright, 2010). The translation of items without careful considerations to the target language and culture may yield inaccurate interpretations and

limited score comparability (Oliveri, Ercikan, & Simon, 2015). Translated and original versions of a test may not be psychometrically equivalent. Double monolingualism does not provide a satisfactory solution from either a theoretical or practical standpoint.

1.4 Assessing the Content Knowledge of Emerging Bilinguals

The difficulty of assessing emerging bilinguals is more pronounced for content assessments, or assessments that measure a construct that is not language proficiency. In addition to all the complexities of test design and development, assessment specialists need to consider the relationship between the proficiency in the different languages and content (García, 2009). Language proficiency could be treated as part of the *construct* measured by the assessment, or not. The construct of a test is the concept that the test is designed to measure (AERA, APA, & NCME, 2014), the attribute assumed to be reflected in test performance (Cronbach, & Meel, 1955). A person may possess the attribute, or not, or possess some degree of it. In every case, constructs are not observable. Because they are not observable, yet intended to be measured, test developers need to operationally define what is conceptually considered part of the construct and what is left out. As the Standards for Educational and Psychological Testing (AERA, APA, & NCME, 2014) specify, to support test development, the aspects of the construct that are to be represented by an assessment need to be clear. In the context of content assessments for emerging bilinguals, such clarity demands a thorough understanding of the relationship between language - or language practices, if we are to move forward - and content proficiency. Moreover, it requires understanding the relationship between the development of linguistic practices and the acquisition of content. The construct of a content assessment for emerging bilinguals may consider language practices, and such a decision relies on theory and the views of

those involved in the assessment process. However, there is some consensus within the psychometric community that content assessments for emerging bilinguals should remove those language elements that are not part of the construct intended to be measured.

1.4.1 A Traditional Solution: Linguistic Accommodations

A traditional solution to reduce unnecessary linguistic complexity in content assessments has been to use *linguistic accommodations*. The term *accommodation* refers to the adjustments to materials or procedures to increase the accessibility of students who otherwise may not have the possibility to demonstrate their knowledge or skills as intended (Thurlow & Kopriva, 2015). Accommodations encompass all changes to materials or procedures that do not change the focal construct. In particular, linguistic accommodations are those that intend to reduce the linguistic barriers that content assessments may create for emerging bilinguals (Sireci & Faulkner-Bond, 2015). In other words, changes to the test that attempt to eliminate the unnecessary linguistic complexity that may be affecting emerging bilinguals. Following Rivera, Collum, Shafer Wilner, and Sia's (2006) taxonomy, linguistic accommodations can be classified as direct or indirect. Direct linguistic accommodations refer to changes to the language of the test, whereas indirect linguistic accommodations refer to those that change the conditions under which the test is taken, that may help students processing language (Rivera et al. 2006). Direct linguistic accommodations often cited in literature include: provision of a dictionary or glossary in the language of the test, simplified language, provision of bilingual dictionary or glossary, translated items and/or directions, dual language booklet, customized dictionaries, and picture dictionaries (Abedi & Evers, 2013; Kieffer, Lesaux, Rivera, & Francis, 2009; Pennock-Roman & Rivera, 2011; Sireci, Li, & Scarpati, 2003). Indirect linguistic

accommodations often cited in literature include: extra time, small group administration, and test breaks. The purpose of these accommodations is to increase the “accessibility” of emerging bilinguals to content assessments. Accessibility refers to increasing the opportunities for students to demonstrate what they know and can do with regard to the target construct (Kettler, 2015).

Table 1. Linguistic Accommodations by Type

Direct Linguistic Support	Indirect Linguistic Support
1. Directions translated into native language	1. Test time increased
2. Audiotaped directions provided in native language	2. Test schedule extended
3. Written directions provided in native language	3. Subtests flexibly scheduled
4. Oral directions provided in native language	4. Test administered at time of day most beneficial to test taker
5. Directions explained or clarified in native language	5. Breaks during test sessions
6. Test items read aloud in native language	6. Test individually administered
7. Audiotaped test items provided in native language	7. Test administered in small group
8. Language reference materials (mono- or dual-language dictionaries or glossaries) provided	8. Teacher faces test taker
9. Side-by-side bilingual versions of the test provided	9. Test administered in location with minimal distraction
10. Translated version of test direction and/or items provided	10. Test taker provided preferential seating
11. Oral response in native language translated into English	11. Test taker tested in separate location (or carrel)
12. Written response in native language translated into English	12. Special test preparation provided
13. Directions simplified	13. Person familiar to test taker administers test
14. Key words or phrases in directions highlighted	14. ESL or bilingual teacher administers the test
15. Test items read aloud in simplified or sheltered English	15. Additional one-to-one support during test administration in general education classroom (e.g., instructional assistant, special test administration, LEP staff, etc.)
16. Key words and phrases in test highlighted	16. Test administered in familiar room
17. Simplified or sheltered English version of test provided	17. Test administered in ESL or bilingual classroom

18. Directions read aloud in English
19. Directions repeated in English
20. Audiotaped directions provided in English
21. Both oral and written directions in English provided
22. Test items read aloud in English
23. Audiotaped test items provided in English
24. Directions explained or clarified in English
25. Test taker verifies understanding of directions
26. Words on test clarified (e.g., words defined, explained)
27. Spelling assistance, spelling dictionaries, spell or grammar check

From Rivera et al. (2006)

Linguistic accommodations in content assessments operate under two assumptions. The first assumption is that there is a redundant linguistic complexity that affects emerging bilinguals and that can be simplified. In psychometric terms this means that there is construct-irrelevant variance, or variance that is related to characteristics other than the focal construct (Messick, 1989). But construct irrelevant variance can and should be removed early on: under this assumption, linguistic accommodations are only solving a problem that can be resolved with better test development (Abedi, & Linqunti, 2012). The second assumption behind these linguistic accommodation policies, at least in U.S. large-scale assessments, is that they are needed on a temporary basis (Abedi, 2004, 2009). That is, there is an expectation that after some time, emerging bilinguals will catch up with their native peers. Ideologically speaking, accommodations are congruent with weak forms of bilingual education, where the expectation is an eventual language shift. However, such expectations may not be pedagogically appropriate. Research shows that it takes a long time for second language learners to achieve the language proficiency of native students (Abedi & Gándara, 2006; Cummins, 2000, Shohamy 2006). This time is even longer in

the context of content areas such as mathematics (Shohamy, 2006; 2011) as it takes longer to achieve mathematics proficiency because such achievement requires mastering both content and a second language. Also, there is evidence that students continue to use their first languages for a long time in academic situations (Shohamy, 2006, 2011). Linguistic accommodations are based on expectations of eventual or double monolingualism. But bilinguals do not behave as two monolinguals (García, 2009; Shohamy, 2011), so linguistic accommodations are not satisfactory as solutions to the language barriers that bilinguals face in content assessments. Moreover, all these distinctions are based on a fictitious notion of enumerability of languages.

1.4.2 An Alternative Framework to Imagine Assessments

A different approach to address the challenges of developing and implementing fair assessments for emerging bilinguals, is rooted in the concept of *translanguaging*. Translanguaging is a term originally coined by Cen Williams that refers to the language education pedagogy in which students' receptive language is different to their productive language (García, 2007). In this work, we refer to translanguaging as the flexible use of linguistic resources that characterizes bilinguals in their attempt to make sense of their bilingual worlds (Creese & Blackledge, 2010; García, 2009). Translanguaging is a linguistic practice that characterizes bilinguals.

Translanguaging is also considered a lens, an approach to bilingualism (García, 2009). As such, the salient features of translanguaging are (a) an integrated view of languages, and (b) a conceptualization of language as a social practice. Translanguaging as a lens does not conceive the languages of bilinguals as separate but treats them as one integrated system (Canagarajah, 2011; Velasco & García, 2014). Translanguaging assumes

that what makes human communication unique is selecting features from the entire linguistic repertoire to produce an intended message. The focus of translanguaging is not on “languages” but in the way in which bilinguals use their linguistic resources. In addition, translanguaging as a lens focuses on the social aspects of language. Bilinguals engage in translanguaging to negotiate meaning, to perform identities, to convey information, to express themselves, among other functions (Creese & Blackledge, 2010; García, 2009; Velasco & García, 2014). A translingual approach to bilingualism is explicitly concerned about these functions that linguistic practices accomplish.

The concept of translanguaging is related to but it is not the same as *heteroglossia*. Heteroglossia is a theoretical lens that conceives discourse as full of diversity: heteroglossia refers to the diversity in speechness, in languageness, and in voicedness (Blackledge & Creese, 2014). Heteroglossic views on discourse allows us to see that there is a lot going on in a normal conversation: people are performing identities, indexing their views upon certain matters, etc. The concept of heteroglossia is therefore, related to the concept of translanguaging. In particular, an heteroglossic approach to bilingualism also views languages as interrelated (García, 2009). However, heteroglossia is:

Not only – in fact not principally – about the simultaneous use of ‘languages’, but rather refers to the coexistence of different competing ideological points of view, whether constituted in a single national ‘language’ (as Bakhtin proposed) or within the complex communicative repertoires in play in late modern societies (Blackledge & Creese, 2014, p.5).

To view language as an heteroglossic practice is not the same as viewing it as a translingual practice. Translanguaging is about normalizing bilingualism without diglossic separation

(Blackledge & Creese, 2014). Heteroglossia admits diversity, even within what we typically call a language. Both translanguaging and heteroglossia share essential similarities, yet the terms are not equivalent.

Translanguaging provides an alternative framework to examine and solve the problems that content assessments for emerging bilinguals possess. In particular, translanguaging offers a framework to rethink the assessments of bilinguals. Translanguaging is supported on an empirical basis, as it reflects common behaviors encountered in the bilingual classroom (Canagarajah, 2011; Creese & Blackledge, 2010; García 2007;). In this sense, a translingual framework would enhance the alignment between assessment and instruction. Notwithstanding, the paramount advantage of using a translingual when thinking about assessments, is that translanguaging departs from the notion of languages as structured, fixed, and standardized systems of signs. Translanguaging overcomes the epistemological problem of treating language as an autonomous system, attuned with the disinvention project that Makoni and Pennycook (2007) call for. Translanguaging is certainly a promising approach to reimagine the assessments of bilinguals.

1.5 Statement of Purpose

Traditional content assessments for emerging bilinguals are problematic in multiple ways. First, these assessments are typically monolingual, departing from the practices of bilingual classrooms. Bilingual students and their teachers engage in linguistic practices that are unique to themselves, and that involve the creative use of their linguistic resources to produce meaning. These assessments fail to recognize the wide variety of linguistic practices that test takers bring. Second, monolingual content assessments are questioned

on ideological grounds, as they promote or enforce the use of one standardized version of a language, attempting against language rights and diversity. This is questionable and fairly inappropriate in a world where multilingualism is the norm. Third, a more serious problem with monolingual content assessments is that they cannot properly measure the proficiency of emerging bilinguals. These students are acquiring language at the same time that they are acquiring content, and are likely to face higher cognitive loads. Students may master content but are unable to properly demonstrate it because of language barriers.

Importantly, most of the current criticism towards assessments for emerging bilinguals fails to acknowledge the invented nature of language. The problems and solutions are framed in terms of language diversity, linguistic rights, mother tongues, monolingualism, or multilingualism, all of which are built on the conception of languages as entities, as autonomous systems that exist beyond ourselves and that are discernable. More powerful criticism towards these assessments should be able to escape the trap embedded in the conventional concept of language. As long as we are unable to re-conceptualize what we understand by language and languages, all of the solutions that we come up with may fail similarly. The challenges of assessing the content knowledge of emerging bilinguals should be defined in new terms. Translanguaging as a lens and as a linguistic practice provides a cornerstone for developing and implementing fair assessments for emerging bilinguals.

Using translanguaging to develop and implement content assessments requires normalizing bilingual competence (García, 2009). Concretely, it requires incorporating linguistic flexibility at the level of questions and responses, both in terms of language and in terms of modalities. This is conceptually different than using direct linguistic

accommodations, because the flexibility that translanguaging points to, is not based on any assumption of redundant linguistic complexity, nor build under any expectations of current or eventual double monolingualism. Solutions based on monolingual expectations will never tap the critical characteristic that emerging bilinguals possess: a flexible, fluid and strategic use of multilingual resources (Canagarajah, 2011; Franceschini, 2011; García, 2009; Velasco & García, 2014). To properly accommodate the practices of every student, the linguistic flexibility that we are talking about has to be self-regulated and dynamic. Students need to be able to activate their entire set of resources, and deploy them according to their individual preferences. We do not need to provide students with “access” to the content of a test, but rather create and implement tests that are built after the language practices that bilinguals use to make meaning and to communicate.

In addition, a translingual approach to assessments requires incorporating opportunities for students to interact with others (López et al., 2017); translanguaging takes place in social interactions. If assessments are to follow the language practices of students, assessments for bilinguals should recreate the conditions in which translingual practices occur, including some degree of interactivity. Therefore, developing and implementing tests using a translanguaging lens, goes beyond merely incorporating linguistic flexibility. Translanguaging entails creating tasks and test in a totally different manner. It is slightly easier to imagine this for language proficiency tests, which would need to assess proficiency in terms of the ability to use language creatively and to produce meaning in multiple contexts and modalities (García, 2009; Hornberger, Lu, Jones, Royster, & Trimbur, 2011; Pennycook, 2008). For example, a task in a language proficiency tests could be to ask a pair of students from different backgrounds to analyze a conversation and

figure out what the third person was trying to accomplish. However, the task of imagining new content assessments is a little bit more complicated, as the inventions and/or conventions of each subject are less negotiable. Developing and implementing translingual assessments requires a lot more deliberation and experimentation.

A starting point to move towards a fully translingual framework, is to incorporate more linguistic flexibility in the assessments of bilinguals. But incorporating more flexibility carries some important challenges. For example, it creates great challenges for scoring assessments. Language use in content assessments would be evaluated holistically, focusing on the overall product and meaning produced by the student, instead of focusing on a particular use of grammatical and/or syntactic features. But scoring items in this way, would necessarily require scorers who are proficient in the multiple languages of the test takers (López et al., 2017) and who master the content of the assessments. This is a limiting restriction in many educational contexts. In light of the challenges that translanguaging entails, what types of tasks should we create? Or how can we incorporate flexibility and yet produce comparable results? Is comparability relevant at all under a translingual approach to test development? Plenty of questions emerge, and these questions require a lot more deliberation and investigation to be properly answered.

As for now, we know that a translingual approach should increase the opportunities of emerging bilinguals to demonstrate their ability to use content across different subjects. But we are far from having an imagery of how the new assessments should look like. Therefore, this work attempts to move the discussion one step towards that direction, by gaining further insight on how the implementation of translanguaging in content assessments should look like. The purpose of this study was to evaluate the effectiveness

and appropriateness of implementing a translingual administration of a mathematics assessment of bilingual girls in the Democratic Republic of the Congo. By translingual administration I mean an administration that incorporated linguistic flexibility at the level of instructions, questions, and answers. By translingual administration I also mean a test that incorporated an element of interactivity. The emphasis on the word administration is to state that test was not developed using a translingual framework: it is only the administration that was re-conceptualized accordingly.

CHAPTER 2

LITERATURE REVIEW

2.1. The Democratic Republic of the Congo

The Democratic Republic of the Congo (DRC) is the largest African country in terms of area of coverage (de Saint Moulin & Tshibanda, 2005). Located in Central Africa, it has a population of about 77 million people (Central Intelligence Agency, 2015). Geographically, its population is dispersed into its ten administrative provinces and one city-province (Kinshasa). The reality across provinces varies importantly, a phenomenon that is explained by the different climates and also, a rich diversity in ethnicities and cultures across the country. Ethnic groups and corresponding cultures are distributed in a way that in different areas of the countries, people may speak languages that are mutually intelligible. DRC is a complex country from a social, cultural, and geographical standpoint. It is geographically and symbolically, the heart of Africa (Trapido, 2015).

DRC is a relatively young country, which gained its independence in 1960 after being officially a Belgian colony for 52 years. The early independence years were characterized by political and social stability. This stability ended in a *coup d'état* seized by J. MOBUTU, who stayed in power for 32 years and renamed the country Zaire. After several conflicts in neighboring countries, a massive inflow of refugees, and a consequent civil war in the East side of the country, MOBUTU's government was toppled in 1999. The leader of the rebellion was named the new president. After two years in power, Laurent D. KABILA was assassinated in 2001, and his son, Joseph KABILA was named the president of DRC. He installed a transitional government and was democratically elected

in 2006. Joseph KABILA is the current president of DRC and the situation in the country has been mostly calm, yet still restless because of the insurrection of some ethnic and/or political groups in certain parts of the country (CIA, 2015; Edinga, 1999).

2.1.1 Linguistic Debates in the Democratic Republic of the Congo

Contemporary language debates in DRC can be framed in the context of larger Sub-Saharan Africa. The language debates in this region are still contested and unsettled. Most of the countries in the region recognize similar (post-colonial) challenges in terms of language planning and policy: the extinction or endangerment of minority-languages, a devaluation of national languages or mother tongues in public spheres, and poorly implemented solutions to the dilemma of how to educate children with respect for their identities and cultures, while preparing them to participate in modern life (Zsiga, Tlale Boyer, & Kramer, 2014). These language debates are based on the idea that there are three main categories of languages: *vernaculars*, *national languages*, and *international languages*. Vernaculars would refer to the language variety that is used in everyday conversations by a community of speakers, and which are different from the standard varieties of the same language, which would only be used in formal and public occasions (Hudson, 1996). National languages would refer to *indigenous* or local languages that are widely used and potentially known to a large group of the population (Zsiga et al., 2014). Indigenous languages would refer to those languages that existed prior to colonialism (Makoni, & Pennycook, 2007), that are local or native. And international languages would refer to those foreign languages such as English and French that were brought to Africa via colonialism (Zsiga et al., 2014). Most of contemporary mainstream language debates in

this part of the world, including those concerning education (e.g. UNESCO, 2007), seem to be framed in these terms.

The vernaculars, lingua francae, and international languages are the essential make-up of language debates in DRC since colonial times. As Meeuwis (1999) explains, the first relevant ideological language debates during colonial times around whether to choose French or local languages as the language of “civilization”. Among the first linguists in DRC we find the protestant and catholic missionaries who arrived to the area with the purpose to “civilize” and evangelize local communities (Makoni & Mashiri, 2007; Meeuwis, 1999). Many of the missionaries were Belgian or French and so, the issue of which language to use in their missions, their own (French) or the ones they encountered (different languages in different settlements) was a central one. It is fairly easy to re-frame this debate in terms of “local” versus “international” languages.

One of the most relevant language debates that took place during colonial times though, was around which language to use for civilization, local vernaculars or lingua francae (trade language). The rapid expansion of missions in DRC, fostered by the liberal boundaries of King Leopold II, brought a large number of Flemish priests to the region. Some of the most influential missionary-linguists of that time were Flemish nationalists: Vyncke, Van Henckthoven, and Hulstaert. The Flemish nationalism of these group of linguists should not be overlooked, as these groups believed that back home, the French language and culture had corrupted the national and natural character of Flanders, post Belgian independence. These missionaries were active members of groups that opposed the imposition of French and French culture in the lives of Flemish people, and their approach to linguistic problems encountered in DRC, was tainted by this membership. To

these missionaries, the issue was not on whether to use French or local languages: using French was out of discussion. The most notorious debate among these Flemish missionaries was on the issue of whether to use “true” local languages or *linguae francae*, which were deemed more appropriate to achieve their purposes (Meeuwis, 1999).

Those missionaries who defended *lingua francae* did so based on the idea of unity, on the possibility of using one single or few languages to accomplish their goals. Using vernaculars would never result in such an outcome. In particular, some proposed to use Lingala – a *lingua francae* – as the medium to civilize and educate Africans, because it had widely spread and adopted in various regions. However, even those who favoured languages as Lingala as the instrument of colonization, recognized that these languages were structurally poor and culturally detached: in response, they began producing dictionaries, grammars, and all the devices that colonials used to invent language in Africa (Makoni & Mashiri, 2007). On the contrary, those missionaries who favoured the use vernaculars, did so based on the argument that any foreign language would corrupt the natural soul of the Africans. To these proponents “Understanding the ‘depths of the African’s soul’ as well as liberating Africans from pre-Christian darkness were both unrealizable without knowledge and use of the Africans’ own language” (Meeuwis, 1999, p.399). This group argued that any corrections to Lingala or similar languages would fail in the same way that French would fail: upgraded *lingua francae* were also foreign. Meeuwis (1999) exposes how these debates were influential at the structural level and with regards to the reach of certain languages, such as the Lingala.

Importantly, this author explains that the debate between vernaculars and *lingua francae* faded away in postcolonial times. In Mobutu’s Zaire, the core discussion

paralleled discourses of anti-imperialism and anti-colonialism, with the real opposition being between African languages and French (Meeuwis, 1999). All African languages, regarded their status of vernacular or lingua francae, were bundled together and contrasted to French, the language of the colonialists. Some Zairean intellectuals wanted to get rid of French from Congolese society, and others wanted to articulate a more radical linguistic program, but the discussion was always framed in terms of French vs. the African languages (Meeuwis, 1999). We can state that this situation has not fundamentally changed. Contemporary language debates in DRC are framed in terms of French, and four lingua francas typically referred to as *national languages*: Lingala, Ciluba, Swahili, and Kikongo (Bokamba, 2009; Edinga, 1999; Meeuwis, 1999). This is similar to the African model, where language policies are established in terms of whether they opt for a colonial language, for an ex-colonial language, or for a combination of both (Makoni, & Mashiri, 2007).

It seems important to highlight how these contemporary linguistic debates are built after the idea of language as an autonomous system. This, despite that there is clear evidence on how the African languages were constructed, even those considered lingua franca, even those considered indigenous or mother tongues. Languages such as Lingala were considered linguistically poor, and were consequently redefined by missionaries, who created grammars, dictionaries, syntaxes, so that these languages became apt to serve the evangelization purpose (Makoni & Mashiri, 2007; Meeuwis, 2009). On the other hand, indigenous languages were a post-colonial creation to recreate an imaginary pre-colonial Africa; they were a response to colonialism (Makoni & Mashiri, 2007; Makoni & Pennycook, 2007). Today, some conceive mother tongues or indigenous languages as more

authentic, but by doing so, they fail to realize that a lot of that authenticity was lost in the process of their creation. Indeed, one of the biggest problems with the invention of these languages, was that they forgot to account for some mixed varieties, some vehicular languages, and Creoles (Makoni & Mashiri, 2007). The issue was that Europeans were interested in creating ideal languages, not just languages (Makoni & Mashiri, 2007), so in their process they decided what to leave in, and what to leave out. The linguistic debate in DRC remains colonial.

2.1.2 Educational System in the Democratic Republic of the Congo

Researchers distinguish three major stages in the evolution of the Congolese educational system (Edinga, 1999). First there is the *traditional stage*, before colonization. In this stage, education was mostly non-formal and concerned with daily life issues. The central objective of education was to integrate individuals to their clans, to render them respectful towards the traditions, and to teach them how to protect and reproduce the group structure in which they were inserted (Edinga, 1999). Second, there is the *colonial stage* that runs between 1885 and 1960, when Congo was either under the ruling of King Leopold II (until 1908) or of Belgium. Colonial education served the ideologies and economic interests of the King and of Belgium. Issues such as the mechanisation of labor or the need of adapting rapidly to the technological changes, were prominent during this stage (Edinga, 1999). Last, there is the *post-colonial stage*, after Congo gained its independence from Belgium (Edinga, 1999).

The post-colonial stage of educational system represents the modern phase of education in DRC. Initially and after the attainment of independent, the main goal of the system was to increase democracy in education (Edinga, 1999). However, after the rise of

Mobutu, that goal evolved towards a more traditional view. Mobutu was primarily concerned about the issue of *authenticity*, which refers to the right of being oneself or the right of being who we are without being necessarily what others want us to be (Edinga, 1999). Mobutu's idea of education was consequently, more concerned with the idea of initiating students into the traditional Congolese life than anything else. In line, the Government established that education had the purpose of forming harmonious Congolese men and women, responsible citizens, citizens who can serve society and promote the development of the country and its culture (UNESCO-IBD, 2010). During Mobutu's Zaire, schools had to form productive, creative, conscientious, cult, free and responsible citizens, open to social, cultural, aesthetic and spiritual values (UNESCO-IBD, 2010). This vision around education – more concerned with the cultural, moral, and affective components of life– brought certain changes to the system.

Some of the modifications under Mobutu's regime were directed towards a revalorization of local languages in formal education. The former educational program (from 1963) was considered the worst in terms of language policy: among other things, it banned the use of indigenous languages in primary education (Nthawakuderwa, 1985). Under Mobutu's regime, this policy was partly reversed. In particular, the educational program introduced in 1974 reintroduced the 4 *linguae francae* as a medium of instruction between 1st and 4th grade (Nthawakuderwa, 1985). The program, however, retained French as a medium from upper primary (5th grade) up to secondary and university levels (Nthawakuderwa, 1985). This practice has not fundamentally changed since then. Currently, these same four languages (a.k.a. national languages) are the language of instruction in grade levels 1 and 2, and French is the language of instruction from grade

level 3 upwards (L'ASBL Investing In People (IIP), 2014). However, in practice, the implementation of this model of bilingualism is largely unstandardized.

In terms of structure, from 1977 on, the educational system in the Congo recognizes two types of schools: public schools and private schools (Mukala-Missumbi, 2012). Private schools do not receive public funds. Public schools are themselves divided into two types: *conventionnees*, or administrated by religious organizations or military organizations, and *non-conventionnees*, or administered by the State. These schools provide primary and secondary education to Congolese students, with most public schools being *conventionnees* (L'ASBL Investing In People (IIP), 2014). In addition, since 1986, primary schooling became mandatory. In line with this view, the current constitution of 2006 established that primary school is a right and so public schools are free of charge. These decisions reflect the political will of providing education for all. However, in practice, schools are not free and parents have to pay small fees.

Currently, the educational system depends upon three Ministries. First, a Ministry of Primary, Secondary, and Professional Education (*Ministere de l'enseignement primaire, secondaire et professionnel*). Second, a Ministry of Higher Education (*Ministere de l'enseignement superieur et universitaire*). Third, a Ministry of Social Affairs, Humanitary Action, and National Solidarity (*Ministere des affaires sociales, action humanitaire et solidarite nationale*) (Mukala-Missumbi, 2012; UNESCO-IBE, 2010). The system operates under the following principles: (a) every person possesses the right of education, (as per the article 43 of the 2006 Constitution), and (b) no Congolese may be discriminated in what refers to education based on religion, family origin, social condition, residence, opinions or political convictions, race, ethnicity, tribal membership, cultural or linguistic

minority (as per the article 13 of the 2006 Constitution) (Mukala-Missumbi, 2012). The paramount goal of contemporary Congolese education is to educate citizens who are competent, who share core human values, and who work towards the achievement of a democratic, solidary, prosperous and pacific society (UNESCO-IBE, 2010).

2.1.3 The Structure of the Educational System

The Congolese educational system distinguishes four cycles: *école maternelle*, *école primaire*, *école secondaire*, and *école supérieur*. First, preprimary schooling (*école maternelle*) is mandatory, and organized into one cycle of three years. Second, primary schooling (*école primaire*) is also mandatory and organized into one cycle of six years. These years are split into three two-year degrees: elementary, medium, and terminal degrees. At the end of primary schooling, students need to take a standardized test to obtain the certificate of completion (test national de fin d'études primaire (TENAFEP)). Third, secondary school (*école secondaire*) is divided into two cycles: a short and a long cycle. Different programs are offered in secondary schooling, and the duration of the cycles depend on the type of secondary education. For example, schools of arts and crafts have a duration of three years, but general schooling cycle has a duration of six years. The end of secondary schooling is also sanctioned by the approval of a national standardized test called *Examen D'Etat*. Last, tertiary education (*école supérieur*) distinguishes technical and pedagogical institutes from universities. Each type of institution offers programs of different nature (vocational, teaching, or professional) and has its own cycles (Mukala-Missumbi, 2012; UNESCO-IBE, 2010).

2.1.3.1 The Characteristics of Primary School

In DRC, primary schooling lasts 6 years, from first to sixth grade. The theoretical age for this stage is between 6 and 11 years (UNESCO-IBE, 2010). As a rule, no student older than 9 years old may be admitted to the first year of primary schooling (UNESCO-IBE, 2010). However, because of the large repetition rates (up to 27% in first grade) and of attrition (only 44% of students who enter primary schooling make it through 6th grade), students enrolled in primary school have a wider age range. Indeed, recent data shows that students between 3rd and 6th grades span a wider age range, from 7 to 21 years old (Randall, 2015).

The main purpose of primary education is to prepare children to insert themselves in society and continue their studies (MEPSP, 2009; UNESCO-IEB, 2010). Academically speaking, after primary schooling, children should possess fundamental knowledge of languages and of mathematics. To that end, the curriculum comprises 16 subjects grouped in three: (a) three subjects that provide instrumental knowledge and skills (e.g. languages), (b) six subjects that provide other scientific knowledge and skills (e.g. education for health and the environment), and (c) seven subjects that comprises aesthetic activities (e.g. music) (MEPSP, 2009; UNESCO-IEB, 2010). These activities demand around 30 hours a week, with a higher proportion of time spent in the first group of subjects.

2.1.3.2 Mathematics Instruction in Primary School

Mathematics is one of the core subjects in Congolese education. It is formally conceived within the first group of subjects recognized in the curriculum (subjects that provide instrumental knowledge and skills). The number of hours per week for

mathematics instruction goes from 5 between grade levels 1 and 4, to 7 between grade levels 5 and 6. This corresponds to the second highest number of hours per subject, after French (MEPSP, 2009). These characteristics have been stable for a long time, but they could change in the near future. Recently, UNICEF conducted an evaluation of mathematics instruction in Congolese primary schools (L'ASBL Investing In People (IIP), 2014). The appointed group evaluated in depth the math instructional practices, materials, standards, evaluation practices, and students' mastery of the subject. The analysis exposed a series of problems with regards to mathematics instruction, followed by an important number of recommendations, including a new set of standards. The expectation is that the mathematics curriculum and instruction undergo some changes in the near future.

Currently, the curriculum for mathematics is organized as a three-layers set of objectives. First, there are two general objectives that apply to the whole cycle. The general objectives for mathematics are that students are: (a) able to solve daily life problems that require mathematical concepts, and (b) able to approach new situations in order to find solutions [*“aborder une situation nouvelle (se poser des questions, s’informer, rechercher par soi-même, réfléchir, développer le sens pratique, ...) pour trouver des solutions.”*]. Second, there are fifteen intermediate objectives. These intermediate objectives apply for two-year cycles (1st and 2nd grade, 3rd and 4th grade, 5th and 6th grade). Intermediate objectives are organized by competence: (a) numeration, (b) operations, (c) “grandeur”/sizes/quantities, (d) geometric shapes, and (e) mathematical problems. The current (official) intermediate objectives are presented in Table 1. Last, there are more than two-hundred specific objectives for each academic year, organized under their corresponding competence (MEPSP, 2009; UNESCO-IEB, 2010).

The competencies are described in more depth in each specific objective. In primary education, there are 277 specific objectives for mathematics, which go from 37 in grade

Table 2. Intermediate Objectives for Mathematics Instruction in Primary Schooling

Degree Elementaire	Degree Moyen	Degree Terminal
Count, name and compare mathematical objects, read and write using numbers and letters numbers between 0 and 100	Count, name and compare mathematical objects, read and write using numbers and letters numbers between 0 and 10 thousand	Compose, name and compare mathematical objects, read and write using numbers and letters large numbers
Add, subtract, multiply, and divide simple numbers. Compute doubles, halves, quarters, triples, and thirds of numbers between 0 and 100	Perform mentally and written the fundamental operations with integers, decimals, and simple fractions	Perform mentally and written the fundamental operations with decimals and fractions
Compare, measure, estimate, and evaluate lengths, masses, and capacities. Being able to communicate times and to use national currency	Measure, estimate, evaluate, weight, verify and pay, in relation to length, capacity, mass, time, and money measures	Establish the relationships between different measures of size
Situate, orientate oneself in a space, re-draw objects. Bend, cut, and draw geometric shapes. Draw lines with or without the aid of rulers	Identify and build geometrical figures and bodies, calculate their perimeter, dimensions and area	Identify and build geometrical figures and bodies, calculate their dimensions, area, and volume
Solve problems with simple directions	Solve simple problems	Solve complex problems

Source: MEPSP, 2009.

level 1 to 58 in grade level 6 (L'ASBL Investing In People (IIP), 2014). An example of these objectives is “to calculate the buying price, the selling price, the revenue and loss” (MEPSP, 2009). A close analysis of the specific objectives indicates that the Congolese educational system is not covering certain skills deemed essential to the development of numeracy. In particular, the current curriculum does not cover algebraic conscience nor some statistical analyses commonly found in the curricula of other countries (L'ASBL Investing In People (IIP), 2014). These and other problems should be addressed in an upcoming renewed program for mathematics curricula.

In terms of language, mathematics instruction follows the general Congolese policy. The language policy for instruction in DRC is such that it converges to French. In particular, for the first two years of instruction, national languages (Ciluba, Lingala, Swahili, or Kikongo) should be used to ensure learning. From third grade on, French is the main language of instruction for all subjects, including mathematics (UNESCO-IEB, 2010). Despite this Federal mandate, the authorities recognize the need of using multiple languages in instruction in primary school and to provide a smoother transition. It is a common practice in the classroom to use the local languages to ensure that children understand the content being taught, even beyond 3rd grade.

2.2 Assessing Mathematics

Mathematics content is delivered through language (Adoniu & Qing, 2014) and assessment specialists need to become aware that when testing mathematics, they are actually testers of the language of mathematics. A critical task to develop any mathematics assessments, under any approach to test development, is to therefore identify the linguistic demands of mathematics – or the linguistic practices that arise in mathematics contexts – and distinguish them from the linguistic demands or practices that are not relevant to mathematics proficiency. This distinction is central to define what the construct of mathematics proficiency should look like. Conventional research may offer some insight in this regard.

2.2.1 Academic Language as a Framework to Support Assessment Development

In educational and assessment settings, many call for a language model that distinguishes between *conversational* or *social* and *academic aspects* of language

proficiency (Baker, 2011; Cummins 2000). The model draws a line between the highly contextualized “everyday” uses of language, and those uses that pertain to academic settings, and that are often less contextualized and more abstract (Cummins, 2000). An essential difference between the two proposed aspects of language proficiency is that contextualized language is supported by interpersonal and situational cues, whereas decontextualized or academic language relies mainly on linguistic cues to meaning – that is, on the knowledge of the language itself (Cummins, 2000). The strength of the distinction relies in the fact that academic language is a significant predictor of students’ academic success across content areas (Bailey, 2007).

Researchers have shown a renewed interest in the concept of academic language (Bailey & Huang, 2011). Academic language can be described from a vocabulary, a grammar, or a discourse function (DiCerbo, Anstrom, Baker, & Rivera, 2014), and provides a detailed framework to think about language and its acquisition. In particular, the idea of academic language is useful to analyze the linguistic demands of academic disciplines (Bailey & Huang, 2011). In the assessment arena, the distinction between of social and academic language may clarify the relationships that exists between first or second language proficiency, and content knowledge (Cummins, 2010). Not surprisingly, assessments for emerging bilinguals have come to revolve around the concept of academic language (Gottlieb, 2014). The concept of academic language should enlighten the discussion of which characteristics of language are relevant to certain subjects, and therefore, part of the construct to be measured. Because academic language varies across disciplines (DiCerbo et al., 2014), academic language needs to be analyzed from a subject-specific standpoint.

2.2.2 The Academic Language of Mathematics

There is no single view around the language of mathematics, and the relationship between language and mathematics requires further inquiry and discussion (Brown, 2002). Under conventional approaches, it is said that mathematics entails lower linguistic demands relative to other disciplines (e.g. Bailey & Huang, 2011). For example, more passive verb forms are observed in science and social studies than in math, and the sentences used in math are typically shorter and/or grammatically simpler. Yet research shows that mathematics requires higher than expected levels of language command (DiCerbo et al., 2014; Zevenbergen, 2001). Among other challenges, mathematics demands to reason abstractly and quantitatively. To do so, students need to use and master the symbolic, graphic, tabular, formal, and/or technical language commonly used in mathematics (National Governors Association & Council of Chief State School Officers, 2012). Mathematics proficiency requires to be familiar with multiple representations of meaning which adds linguistic challenges. For instance, mathematics' problems are often supported by visuals, which are not always self-explanatory or neutral (Adoniou & Qing, 2014). To correctly understand and respond to math problems, students need to master symbolic, language, and visual literacy. Or for instance, the so called mathematical symbols (e.g. $>$, $<$) may exist in other languages, but with different meanings (Adams, 2003).

The academic language of mathematics entails its own linguistic demands. In math, the structure of sentences may use a different order than that of the language of instruction. Or sentences may rely extensively on dependent and relative clauses, such as sentences asking or describing hypotheses (Adoniou & Qing, 2014). On the other hand, mathematics

carries lexical challenges. Words used in mathematics may have a different meaning than when used in other contexts. For example, the word table can mean “times table” and this *polysemy* can cause difficulties for students, particularly for language learners (Adoniou & Qing, 2014; Martinello, 2008). The language of mathematics evolves across grade levels (NGA & CCSSO, 2012), and the polysemy may increase. Some illustrative concepts from advanced calculus that display this problem include *integration*, *neighborhoods*, *tangents*, *real* or *hyperreal* numbers. The challenge becomes even greater when considering that mathematics uses language in a very precise way, with sentences having little redundancy (Adoniou & Qing, 2014; NGA & CCSSO, 2012). These demands affect all students learning mathematics. Assessing mathematics should be also, and perhaps mainly, about assessing what students can do with the language of mathematics.

It is clear that mathematics entails its own linguistic demands, but it is not clear whether these demands can be disentangled using the social versus academic language that Cummins and other researchers propose. If we are to conceive language in terms of the practices that students, especially bilinguals, engage in to make sense of their world, and therefore, to make sense of mathematics, we cannot exclude social language. The idea on academic language as the central type to think about constructs and assessments, should be replaced by an idea of distinguishing those language practices that are exclusive or most predominant in the mathematics classroom, and those that are not, regardless of their status of social versus academic. Indeed, the social aspect of language is what translanguaging puts forth as the starting point to think about assessments.

2.3 Research on Linguistic Accommodations Relevant to Translanguaging in Assessments

Traditional research shows that mathematics proficiency is related to language proficiency. Research consistently finds that there is a moderate to strong relationship between measures of language and of mathematics, as evidenced by several standardized tests (Abedi & Lord, 2001; Ercikan et al. 2015). In addition, research has found negative correlations between mathematics achievement and the linguistic complexity of the items (Abedi & Lord, 2001). In other words, more difficult items, from a linguistic perspective, are associated to lower levels of performance. In these studies, linguistic complexity is defined by measures that focus on lexical and grammar/syntactic features of items, such as frequency of non-math words, voice of verb phrase, length of nominal, the use of multiple clauses, etc. To some extent, these findings seem obvious: the linguistic complexity of the items should increase the difficulty of the assessment as it imposes additional cognitive loads on students. Indeed, complex linguistic features may slow down the reader, increase the likelihood of misinterpretation, and add to the reader's cognitive load (Abedi & Gandara, 2006). However, there is evidence that some linguistic features may affect emerging bilinguals more than other students. Linguistic complexity is a common characteristic of math items that present differential item functioning against emerging bilinguals (Martiniello, 2008; Wolf & Leon, 2009). Differential item functioning (DIF) occurs when examinees belonging to different groups show different probabilities of answering an item right, at same levels of estimated proficiency (Zumbo, 1999). Wolf and Leon (2009) found that the magnitude of DIF of math items of low difficulty increased with the level of linguistic complexity. In particular, the authors found the largest

associations for items that: (a) presented more academic vocabulary, (b) had higher proportion of language to non-language elements, and (c) were evaluated as “requiring language knowledge” by a panel of experts. Additional analyses of students’ responses show that a reduction of the linguistic complexity of math items may increase the performance of all students, but especially of emerging bilinguals (Abedi, 2006; Abedi & Gandara, 2006; Abedi, Hofstetter, & Lord, 2004; Abedi & Lord, 2001). All these findings suggest that mathematics items in traditional assessments may contain unnecessary linguistic complexity, that is especially detrimental to emerging bilinguals.

It is in this scenario that linguistic accommodations make sense. Linguistic accommodations are meant to solve the issue of construct-irrelevant variance that may arise due to language issues. Bottom line, it is unlikely that non-native speakers perform at their best if they do not understand the questions being asked or cannot respond in the language they are asked to do so (Abedi, 2004, 2006). In response, linguistic accommodations are meant to provide a fair opportunity to non-native speakers to truly demonstrate what they know and can do (Abedi, 2004). Reasonable linguistic accommodations increase access without altering the nature of the construct being measured (Lindstrom, 2010). In particular, good accommodations should result in what is known as the “interaction hypothesis” (Haag, Heppt, & Roppelt, 2015; Sireci, Scarpati, & Li, 2005). On one hand, the interaction hypothesis states that when linguistic accommodations are given to non-native speakers, their test scores should improve relative to the scores that they would attain in the unaccommodated test. On the other hand, the hypothesis states that students without the need for the accommodation should get similar scores in both tests. Meeting these two conditions means that the accommodation effectively minimizes construct-irrelevant

variance without providing unfair advantages – that is, without changing the focal construct. The first condition is meant to assess the effectiveness of the accommodations. The second condition is meant to assess their validity.

The use of proper linguistic accommodations should result in lower score gaps between emerging bilinguals and native speakers (Lindstrom, 2010). However, the effectiveness and validity of an accommodation cannot be taken for granted and needs to be evaluated separately by accommodation. Two of the most used linguistic accommodations correspond to dictionaries and glossaries with which emerging bilinguals can make sense of the critical terms of the test that they may not understand. These types of direct linguistic accommodations are typically used in a paper and pencil format, although current technology enables computer-based alternatives. Dictionaries show promise as an accommodation (Sireci et al. 2003), but there are some nuances to take into consideration. Research in the U.S. suggests English dictionaries may benefit students of intermediate English proficiency (Albus, Bielinski, Thurlow, & Liu, 2001). Research also suggests that this accommodation is sensitive to the type of dictionary used. Commercially available dictionaries differ importantly, for example, in the difficulty level of the vocabulary of the definitions (Kopriva, 2000), so it is difficult to pick one that is effective for the pool of examinees. Therefore, some recommend against using commercial or published dictionaries (e.g. Abedi, Courtney, Mirocha, Leon, & Goldber, 2001). They state that it is better to use customized dictionaries for effectiveness and validity considerations (Abedi, 2004). On the other hand, glossaries also show promise as a good accommodation for emerging bilinguals (Sireci et al., 2003). However, some argue against them saying that, when coupled with extra time (another accommodation), all students perform better

and that without extra time, using glossaries may even reduce scores (Abedi, Lord, Hofstetter, & Baker, 2000). Therefore, glossaries may not be effective nor a valid accommodation. Notably, the only exception would be for the case of computer-based glossaries, which are deemed effective without posing validity threats (Abedi, Courtney, & Leon, 2003).

Another group of linguistic accommodations correspond to change to the materials or the presentation of a test, such as translation of instructions or questions. These accommodations are not used as often as other linguistic accommodations (Abedi, 2004). Research suggests that these accommodations may be problematic and only work under specific conditions or with particular groups of students. Also, because of their nature, it is unlikely that they meet the interaction hypothesis. First language speakers may not perform similarly in the original test – written in their home language - and a test translated that is written in a language that they may not even know. Indeed, the translation of instructions does not show conclusive results in terms of its effectiveness (Abedi, 2004). Some argue that this particular accommodation may only work with more time (Miller, Okum, Sinai, & Miller, 1999). In turn, the translation of items may only work if the instructions are also translated (Abedi, 2004), so they should not be considered an accommodation in themselves. Now, the full translation of tests are difficult to produce, and present issues of equivalence even within one language (Abedi, 2004). Besides, translated assessments are only effective and appropriate if the examinees who take it have an appropriate proficiency level in the language of the test, which is not always the case for populations of students.

Another accommodation type that is worth reviewing is *simplified language*. As stated, linguistic complexity may be a source of difficulty and even bias for emerging bilinguals. Simplified language (a.k.a. plain language) refer to modifications to the words and sentences in a test to reduce unnecessary linguistic complexity (Abedi, 2006; Abedi & Gandara, 2006; Abedi, Lord, & Plummer 1997). These modifications are varied and typically attempt to target lexical and syntactic sources of difficulty. For example, lexical modifications include using frequently used words (instead of rarely used words), and syntactic modifications include eliminating passive voice constructions, conditional clauses, or subordinate clauses, among many others. Overall, tests with simplified language also show promise (Abedi, 2004; Sireci et al., 2003) and may potentially narrow the score gaps between native speakers and emerging bilinguals (Abedi, 2006, Abedi & Gandara, 2006; Abedi et al., 2000). Those who argue for the accommodation state that aside from being effective, it does not alter the focal construct; i.e., it is a valid accommodation across multiple content areas (Abedi & Gandara, 2006; Rivera and Stansfield, 2001). However, others argue that the effectiveness of simplified language is contingent to the language proficiency of examinees, mostly helping students of intermediate language proficiency (Haag et al., 2015; Pennock-Roman & Rivera, 2011). Moreover, some researchers find little reason to believe they are an effective accommodation and state that the results are straight inconclusive or non significant (Haag et al., 2015; Kieffer et al, 2009), possibly since there are few studies looking at it. A technical shortcoming with these accommodations, is that their effectiveness depends on the method used to estimate person ability (Haag et al., 2015). Some argue against these accommodations under the idea that

irrelevant linguistic complexity should always be minimal and that such procedure is a characteristic of sound test development (Sireci & Faulkner-Bond, 2015).

Since most research around accommodations is inconclusive, some researchers have attempted to identify common findings by conducting meta-analyses, which allows us to integrate finding across studies. This technique allows us to derive a pooled estimate of the unknown common true effectiveness or validity of a particular accommodations. In the following paragraphs, we review the outcomes of three meta-analyses: (a) Kieffer et al. (2009), (b) Pennock-Roman & Rivera (2011), and Li & Suen (2012). All three meta-analyses are based in the U.S. context. Following the interaction hypothesis framework, three of these studies evaluated the effectiveness of linguistic accommodations, and two of them evaluated the validity of linguistic accommodations. There is an overlap between the studies used in each meta-analysis. From the 24 articles used in the three meta-analyses, seven are common and only 11 articles are used by only one meta-analysis (with seven of them being used by Li & Suen). However, there are differences in the methods used and results obtained by each work.

Kieffer et al. (2009) evaluated the effectiveness of seven accommodation: (a) simplified English, (b) English dictionary or glossary, (c) bilingual dictionary or glossary, (d) extra time, (e) Spanish language test, (f) dual language questions, and (g) dual language booklet. The mean effect size was evaluated for all accommodations together and for each accommodation separately. The mean effect size across all accommodations was not significantly different from zero. And separately, only one accommodation had an overall positive effect on the scores of emerging bilinguals. In particular, the use of English

Table 3. Characteristic of the Selected Meta-Analyses

	Kieffer et al. 2009	Pennock-Roman et al. 2011	Li et al. 2012
Percent of studies using experimental design	73% (8/11)	100% - with 2 studies following a repeated measures design	Unclear
Number of accommodations or categories of accommodations evaluated for effectiveness	7	11 + combinations or variations	6
Number of accommodations evaluated for validity	5	8 + combinations or variations	N.A.
Total number of effect size for effectiveness	38	50	85
Total number of effect size for validity	30	32	N.A.
Effect size measure	Cohen's D and Hedges G (correction to account for publication bias)	Glass's D and Hedges correction (bias of small control groups)	Hedges G
Test of heterogeneity	Yes (Q statistic)	Yes (Q statistic)	No

language dictionaries and glossaries had a small but consistent positive effect that was robust across different types of studies. The effect, measured with Hedge's G, was 0.15 for the fixed effects design and 0.18 for the random effects design. Neither of the other accommodations was deemed effective. Notwithstanding, some findings are worth highlighting. The study suggest that Spanish versions as well as dual language assessments may favor students who receive instruction in that particular language. This relates to what Abedi (2004) highlighted: that the language of the tests need to match the language of instructions to increase the opportunities of emerging bilinguals.

In relation to the analysis of the validity of accommodations, Kieffer et al. (2009) looked at the effect of five accommodations on the performance of native English speakers:

(a) bilingual dictionaries and glossaries, (b) Spanish versions of the assessment, (c) extra time, (d) simplified English, and (e) English dictionaries and glossaries. An overall estimate of the effect size indicated that taken together, these accommodations did not produce significant changes in the scores of English speakers. The individual analysis showed that only the Spanish version of the assessment was related to a significant reduction of scores. The effect, measured with Hedge's G , was -0.87 . Therefore, it is not reasonable to believe that providing these accommodation to emerging bilinguals would result in any form of unfair advantage.

In contrast, Pennock-Roman and Rivera (2011) examined a gamut of accommodations as well as different combinations between them: (a) simplified English (plain English), (b) bilingual glossary, (c) Spanish version, (d) extra time, (e) English dictionary/glossary (P&P), (f) pop-up English glossary, (g) dual language (questions, booklets, read alouds in Spanish, bilingual answers), (h) picture dictionary, (i) pop-up bilingual glossary, (j) read aloud in English, and (k) small groups accommodations. Some variations were provided. In particular, the following accommodations were also provided with extended time: English dictionary/glossary, simplified English, dual language, and bilingual glossary. The results were different to those found in Kieffer et al. (2009). The most important finding was that the effectiveness of the accommodations was not independent of the language proficiency level of the students. In particular, the most effective accommodation for students with low English proficiency and who received instruction in Spanish, was the translation of tests into Spanish. However, this accommodation was not effective for students receiving instruction in English, whose home language was not Spanish, or with higher levels of English proficiency. Indeed, the

most effective accommodation for students with intermediate English proficiency was simplified or plain English. In addition, there were some consistent trends in relation to time conditions. Larger individual effect sizes were observed when time was extended or generous.

Average effect sizes were estimated for those conditions for which enough and consistent information was available. Three accommodations presented statistically significant outcomes: (a) pop-up English glossary ($D=0.29$), English dictionary/glossary with extended time provided to both control and accommodated groups ($D=0.23$), and bilingual glossary ($D=-0.18$). As noted, the use of bilingual dictionaries was associated to a significant loss in achievement. This is not the case when the accommodation is coupled with extended time limits, suggesting that its use requires the provision of more time to deal with the new material. Indeed, this trend applies to all of the accommodations, which provided with extra time, importantly increase the scores of English learners. Aside from English dictionary/glossary, this is true for simplified English ($D=0.11$, n.s.), for dual language tests ($D=0.30$, n.s.), and bilingual glossaries ($D=0.25$). This remains true for English dictionary/glossary when the extra time is only provided to the accommodated group. The authors highlight this finding by stating that “most accommodations did improve the performance of ELLs beyond a trivial level when students were allowed sufficient time to work with the extra printed material provided” (Pennock-Roman & Rivera, 2011, pp. 21–22).

Pennock-Roman and Rivera (2011) also examined the validity of the accommodations across non English learners. In particular, the validity of the following accommodations was examined: (a) simplified English (plain English), (b) bilingual

glossary, (c) Spanish version, (d) extra time, (e) English dictionary/glossary (P&P), (f) pop-up English glossary, (g) dual language (questions, booklets, read alouds in Spanish, bilingual answers), and (h) small groups accommodations. Some accommodations were provided with extended time for both groups: English dictionary/glossary and simplified English; English dictionary/glossary was also coupled with extra time to the accommodated group only. Overall, the effect of accommodations on English speakers was technically zero, either because the effects were not statistically significant or when they were (e.g. plain English, $D=0.64$), the effect was too small to be considered practically significantly. However, the lack of enough studies meant a loss in statistical power: some effect sizes, while not statistically significant, still deserve examination. In particular, the use of dual language ($D=-0.17$) and bilingual glossary ($D=-0.13$) were associated to non-negligible losses in achievement. Because of their small sample sizes, Spanish version and small groups accommodations were not used in the estimation of overall average effect sizes. However, those few studies suggest non-negligible negative effects.

Li and Suen (2012) looked at six categories of accommodations: (a) linguistic simplification, (b) dual language booklets, (c) Spanish version of the test, (d) dictionaries or glossaries, (e) other accommodations – extended oral presentation, small-group testing, visual support, other provisions, and (f) extra time. The meta-analysis did not evaluate the validity of accommodations but only evaluated their effectiveness. However, they used hierarchical linear modelling techniques which allowed them to understand the role that background variables or study characteristics played in the different outcomes. The effect sizes were used as level 1 variables and the following variables were used in the level 2 modelling: ethnicity, grade level, test subject, English proficiency, and accommodation

type. The most important results were that in average, accommodated students scored significantly better than their counterparts (coefficient of 0.157) and that the only level-2 variable that explained differences in effect sizes was the level of English proficiency. In particular, students with low levels of English proficiency benefited the most from accommodations; in average, the effect size was higher by 0.490 for this group of students.

Research on accommodations shows that the assignation of linguistic accommodations is a very complex endeavor, and that we should stay away from the one size fits all mindset (Abedi, 2004). In particular, literature shows that the effectiveness of different accommodations is not independent of the language proficiency of examinees (e.g. Li & Suen, 2012; Pennock-Roman & Rivera, 2011). Some of the accommodations that work well with emerging bilinguals (e.g. translated tests for students who receive instruction in the translated language, and are proficient in the language) will never be valid. And beyond the issues of validity and effectiveness, some accommodations are not widely used by emerging bilinguals, affecting their potential (Wolf, Kim, & Kao, 2012). All of these findings, that call for multiple ways to accommodate different types of students, make us wonder how standardized standardized assessments can be. A better way to solve the issue of language in content assessments, may be to embed linguistic flexibility and multiple discursive practices, from scratch, in the design and development of the assessments.

CHAPTER 3

METHODOLOGY

3.1 Overview

The main goal of this dissertation is to gather evidence on the effectiveness and appropriateness of implementing translanguaging principles in the administration of a mathematics assessment called the Early Grades Mathematics Assessments (EGMA) – which is widely used in the context of international development. To that end, I approached the study using a mixed methods design; I collected quantitative and qualitative data. Qualitative research is concerned with the understanding of human behavior from an actor’s frame of reference, and therefore, it is subjective in nature (Nunan, 1992). The underlying assumption is that there is no objective reality that lives outside the observers and/or participants but that the features of social environments are socially constructed (Gall, Gall, & Borg, 2007). Qualitative research is therefore, inductive in nature, and better suited to understand beliefs and reasons behind behaviors or preferences, as it does not make assumptions of truth prior to the data collection. By contrast, quantitative research is deductive in nature. Quantitative research begins with a theory or hypothesis and seek to gather data in order to evaluate such theory or hypothesis (Nuan, 1992). Quantitative research assumes that there is knowledge beyond the individual and local setting (Gall et al., 2007). This assumption leads to the aggregation of data across individuals to uncover such knowledge or “truth”. Quantitative methods are therefore, better suited to test theories in large scale samples.

Although a distinction should be made with respect to these different approaches to research, neither approach is intrinsically superior to the other. In this dissertation, the aim is to evaluate both the effectiveness and appropriateness of a translingual administration of the EGMA, in the context of Congolese primary schooling. I chose a mixed methods design because answering the research questions required elements from both research traditions. On the one hand, to evaluate the effectiveness of the approach, I used quantitative methods. The EGMA is widely used to evaluate educational outcomes across groups of students, and its administration is not concerned with the subjective experience of the students. The effectiveness of the translingual administration needs to be evaluated at the aggregated level, because this is how the instrument is typically used; thus, quantitative methods are better suited for this purpose. On the other hand, to evaluate the appropriateness of the translingual features I mostly used qualitative methods of inquiry. In this context, appropriateness refers to how students experienced the translingual administration. It also refers to whether stakeholders believe that translanguaging provides a better framework to assess their students. The evaluation of appropriateness deals with individual views; hence, qualitative methods are better for this second purpose.

The research questions that guided this study are the following:

1. Effectiveness:

- 1.1 Does a translingual administration of the EGMA assessment have an impact on math achievement of Congolese girls, as evidenced by a significant difference in their scores when compared to the traditional administration?
- 1.2 Does a translingual administration of the EGMA assessment allow us to obtain more reliable information about Congolese girls' math knowledge,

as evidenced by higher reliability estimates when compared to the those obtained from the traditional administration?

2. Appropriateness:

2.1 Is the translanguing administration of the EGMA assessment appropriate for the context in which this data was collected, as evidenced by:

- a) The characteristics of the context in which the assessments were administered
- b) The level and ways in which girls engaged in flexible bilingualism
- c) The alignment with teachers' practices and beliefs around how to assess math proficiency
- d) The feedback from enumerators who administered the EGMA

3.2 Research Design

3.2.1 The Early Grades Mathematics Assessment

The Early Grades Mathematics Assessment - or EGMA - is an instrument that measures mathematics knowledge and skills that should be acquired in early grades of instruction, in other words, it targets foundational math knowledge and skills. It was developed by RTI international in 2008 and funded by the United States Agency for International Development (USAID). Since then, the instrument has been implemented in more than 14 different countries and eleven languages (Platas, Ketterlin-Gellar, Brombacher, & Stiabkhan, 2014), being the preferred instrument used to evaluate mathematics across USAID-funded projects. Recently, the EGMA underwent a series of modifications after two panels of experts suggested changes (Platas et al., 2014). The

current version of the EGMA is called the *Core EGMA*, and is the version that we used in this study.¹

The EGMA was designed to reliably measure early grade mathematical skills across a wide set of countries (Platas et al. 2014). The instrument focuses on number sense and operations, as these skills are considered fundamental for early grade numeracy. The test developers recommend two uses of the scores: country level diagnosis and growth measurement. With relation to the first, the rationale is that the EGMA taps skills that develop across countries and that are critical to the development of mathematics proficiency (Reubens, 2009). Indeed, the EGMA was designed in a way that teachers could relate items to the national curricula (Reubens, 2009). Therefore, the authors suggest that large-scale administrations of the EGMA may inform curriculum mastery across different countries. As a growth measure, the EGMA would be suitable for program evaluation. Programs may administer a same form or parallel forms of the instrument at the beginning and at the end of a program, to obtain growth measures. These measures are used to evaluate the impact such programs.

3.2.1.1. Characteristics of the Core EGMA

The Core EGMA was simplified to six subtests (tasks) that taken together, can produce a snapshot of children's knowledge of foundational mathematics competencies (Platas et al., 2014). The EGMA does not cover an important number of subdomains that are typically present in primary grades, such as multiplication and division, or fractions and decimals. The rationale for a test with emphasis on numbers and operations was to

¹ From now on, any reference to EGMA is a direct reference to the Core EGMA

include items that tapped the fundamental foundational skills, while ensuring that the test was not too long (Platas et al., 2014). The content included in the EGMA represents a progression of the foundational skills that support proficiency in math, and comprises teachable skills that are common to many national curricula for early grades. The subtests included in the Core EGMA are: (a) number identification, (b) quantity discrimination, (c) missing number, (d) addition (level 1 and level 2), (e) subtraction (level 1 and level 2), and (f) word problems (Platas et al., 2014). These subtests vary in their specifications and number of items. Some of these subtests are timed. For example, addition and subtraction subtests are timed because speediness is considered as part of the fluency conceptualization. The details of each of the subsets are presented in Table 4.

Table 4. Details of the Core EGMA

Sub-test	Measure	Test Specifications	Number of Items	Timed/Untimed
Number identification	The ability to correctly identify numbers.	The first items include numerals 0, 9, and one other single-digit number. The next 12 items consist of two-digit numbers from 10 to 99. The last five items are three-digit numbers	20 items.	Timed (60 seconds).
Number discrimination	The ability to correctly reason about magnitudes.	The first item is a set of one-digit numbers, the next five items are two-digit numbers, and the last four items are three-digit numbers	10 items. Each item consists on a set of two numbers.	Untimed.
Missing number	The ability to correctly recognize number patterns.	The items are such that eight of the items have increasing patterns, and two have decreasing patterns. Items 1,2, and 6 increase by one, and items 3,4, 5 and 8 increase by tens, hundreds, twos, and fives, respectively. Items 7 and 9 decrease by twos and tens, respectively. The last item with numerals within the range of 1-20 increases by fives, but does	10 items. Each item consists on a pattern of numbers where one is missing.	Untimed.

		not begin with a multiple of five.		
Addition level 1	The ability to add accurately and fast.	No sums are greater than 19 and no addends are greater than 10.	20 items.	Timed (60 seconds).
Addition level 2	The ability to add accurately and fast, with larger addends.	No sums are greater than 70. Items are not given to students who score zero in the addition level 1 problems.	5 items of increasing difficulty.	Untimed.
Subtraction level 1	The ability to subtract accurately and fast.	The items are the inverse of the addition problems.	20 items.	Timed (60 seconds).
Subtraction level 2	The ability to subtract accurately and fast, with larger numbers.	The items are the inverse of the addition problems. Items are not given to students who score zero on the subtraction level 1 items.	5 items of increasing difficulty.	Untimed.
Word problems	These ability of interpreting problems and understanding operations required to solve them.	The items represent a different problem type each: change (result unknown), combine, compare, change (start unknown), sharing, and multiplicative.	6 items that increase in difficulty. Some of the items mirror addition and subtraction level 1 items.	Untimed.

The EGMA is an oral assessment and individually administered. The motive is that children at early grades may not be proficient in reading and/or writing, so an oral administration prevents construct-irrelevant variance due to varying proficiencies in reading and/or writing. Second, the EGMA may be administered in paper and pencil or using tablets. These methods of administration could result in different experiences for students, so enumerators and users of the data have to be aware of comparability issues that may arise because of the different administration modes. Last, the EGMA has to be locally adapted to the context it is being used in. The local adaptation refers to the process of

coming up with a final tool to be used in a given context and for a particular purpose. For example, a central part of the adaptation is to select which subtests to include.

One of the recommendations that RTI sets forth with regards to local adaptation, is to administer the questionnaire in the local language. In their words “Although many view language as being unrelated to mathematics, in reality, language is integral to the use and learning of mathematics” (Platas et al., 2014, p. 2). However, the adaptation needs to be standard as “the assessor must say these words [those of the instruction] exactly as they are written in the instruments” (Platas et al., 2014, p. 34). RTI wants to ensure that if the instrument is translated into a local language, all students who take the test in that local language should undergo the same assessment experience. In addition, RTI recommends children to respond in whatever language they choose, provided that the assessors understand these languages. The rationale is to ensure that the test is measuring mathematics’ content, and not language proficiency (Platas et al., 2014). Therefore, RTI recognizes the role that language plays in accessing mathematics content, but there is no questioning to the concept of standardized language.

3.2.2 A Translingual Administration of the EGMA Assessment

López et al. (2017) suggest a two-step framework to implement translanguaging in content assessments. The first step is to ensure that the assessment draws on students’ entire linguistic repertoires, by enabling students to move back and forth between languages and modalities. The second step is to engage students in interactive practices, by creating student-student interactions and/or student-teacher interactions. The rationale for this interactive component has to do with the fact that emerging bilinguals engage in translingual practices during interaction with others. The typical uses of translanguaging

in the classroom include mediating understanding, constructing or co-constructing meaning, and engaging in identity performances, to name a few (García, 2009), all of which occur in interaction with others. López et al. (2017) two-step framework to develop assessments should enable students (a) to draw from their entire repertoires and (b) to engage in the interactions that stimulate the creative and meaningful use of linguistic resources.

The translingual administration of the EGMA followed this two-step approach. It is important to distinguish between the administration and the test itself. The EGMA was not developed using a translingual approach, and the context in which study was developed prevented me from using a different instrument or to modify it further. This study consisted on the implementation a translingual administration of an instrument originally conceptualized and developed under a monoglossic lens. To implement a translingual administration, two enumerators instead of one, administered the test. The enumerators read the general instructions and the item directions in French - official language of instruction - and/or Lingala - the local language (*lingua francae*). Some of the instructions or directions were read aloud in French, some were read aloud in Lingala, and some were read aloud in both. There was no prescribed pattern to use language in instructions, and each administration was unique in this regard. The only standard condition was that all girls were exposed to a mix of languages as they were told what to do. Second, the girls gave their responses in either French, Lingala, or any flexible mix of them. This was explained to them in the instructions and encouraged throughout the assessment. As part of the administration, the girls could stop for questions at any time, in any mix of language.

This EGMA administration (which I refer to as EGMA TL) shares some similarities with RTI's proposed administration principles. RTI's EGMA administration allows students to listen to the questions in the language of their choice, and to respond in the language of their choice; so does the EGMA TL administration. However, their view around languages is completely different than the one inspiring this study: the objective of RTI's proposed administration is to ensure that language is not a barrier in measuring students' mathematics proficiency. As per current documentation, RTI does not consider mixed and/or nonstandard varieties of languages as acceptable. Most likely, RTI's orientation is that *language is a problem* (García, 2009), a barrier that has to be overcome. The development of the EGMA and its proposed administration, follow monoglossic ideologies.

In contrast, the EGMA TL administration was put together under the conviction that *language is a resource* (García, 2009) all the languages of children are considered important and worth cultivating. The EGMA TL honors heteroglossia and its anti-hegemonic stance around languages. As such, we did not prioritize one language over the other by asking the child to choose a single language of administration. In addition, we did not encourage children to stick to one language in order to provide their responses, and we tried to foster translanguaging at all times. Another major difference with RTI's proposed administration of the EGMA, is that the EGMA TL implemented a two-sided (or three-sided) interactive and flexible bilingualism at the item-, task-, and test-level. The EGMA TL departed fundamentally and practically from the traditional way of administering the assessment.

3.2.3 Sample

This study was conducted in the context of an equating study for a DFID²-funded educational intervention called Vas Y' Filles! The sample used in this study was therefore guided and constrained by the demands of that equating study. The Vas Y' Filles project is an educational intervention that was implemented in five Congolese provinces, and targeted girls that attend schools in grades 3 through 6. The objective of the project was to improve their reading and mathematics skills through a series of treatments that included scholarships, teacher trainings, reading programs, and new instructional material. The Vas Y' Filles! intervention was assigned to certain schools across five provinces in DRC (Kasai, Katanga, Equateur, Bandundu, and Province Orientale). The selection of schools followed a cluster-randomized design, where clusters were initially defined in relation to the provinces (five) and the number of subdivisions in each province (between 2 and 9). To ensure equal representation across all five provinces and their subdivisions, a stratified random sampling technique was used (Randall, 2015). Accordingly, the project selected 737 schools in 212 clusters (Randall, 2015). Since the evaluation of the project followed an experimental design, not all schools within a sampling unit (school clusters) were assigned to the treatment. About half of the schools within a cluster were assigned to the project, and half were left as control schools.

3.2.3.1 Sampling Methodology

The evaluation methodology followed by Vas Y' Filles! was one in which control and treatment schooled were sampled and compared: therefore, the data used for

² Department for International Development (DFID) is the UK development agency

evaluations was based on a sample of control and treatment schools. Per cluster, one treatment and one control school corresponded to evaluation schools (Randall, 2015). However, for methodological reasons, complementary studies had to be conducted in the non-evaluation schools. In line, the equating study took place in non-evaluation schools, and it is among these schools, that I sampled participants.

Specifically, I randomly sampled girls from randomly selected non-evaluation schools. While the equating study took place in several provinces, logistically, it was easier to work in only one of them, as DRC is a country in which moving across regions can be extremely complicated. Therefore, I only used schools from one province: the data collection procedure for this dissertation was carried out between May 11th and June 6th 2015, in the Congolese province of Equateur. During this period, I sampled girls from four schools, where me and two enumerators administered the traditional EGMA (EGMA) to 71 girls, and the translingual administration of the EGMA (EGMA TL) to 80 girls. The girls had self-reported ages between 8 and 15 years, with the age distribution being even across those girls who took the EGMA and those who took the EGMA TL (see Figure 1). In terms of grade level, girls were sampled from grade levels 3 through 6, the distribution being even across grade levels with approximately a quarter of the girls sampled from each grade level. However, the distribution of girls across grade levels varied slightly between girls who took the EGMA and those who took the EGMA TL, as shown in Table 5.

For logistic reasons, the sample of teachers for the interviews was collected from the same schools and grades from where girls were selected. A total of 8 elementary teachers were interviewed. Similarly, the interviewed enumerators were the same as those

Figure 1. Age Distribution of Girls

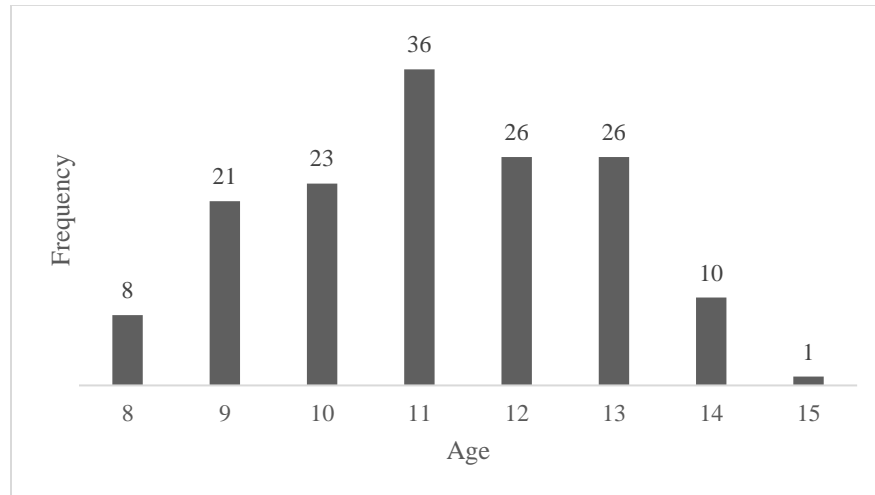


Table 5. Grade Level Distribution of the Girls

Type of Administration	3	4	5	6	Total
EGMA	15	21	10	25	71
EGMA TL	22	13	33	12	80
Total	37	34	43	37	151

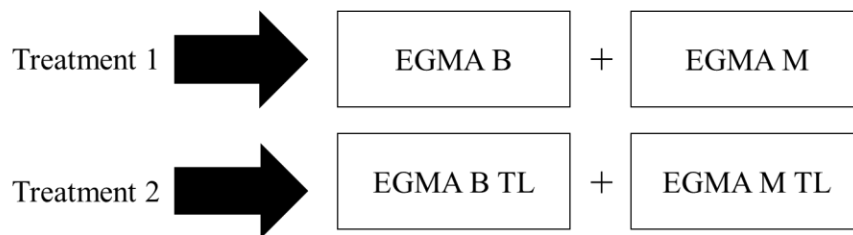
collecting data in these schools, both hired and trained by the project. The enumerators were trained in the EGMA administration and were familiar with the entire process of data collection. They were from the area, and therefore, fluent in the local language (Lingala) as well as in French. I personally trained the enumerators on how to administer the EGMA TL, before the data collection period started.

3.2.4 Procedures

To answer the first two research questions, I administered the EGMA and EGMA TL using a *randomly-equivalent groups design*, where girls were randomly assigned to two different treatments. The first treatment consisted on taking two different forms administered the traditional way (EGMA). The second treatment consisted of taking the

same two forms under the translingual administration (EGMA TL). The different forms were coded B and M since they had been used for baseline and midline data collection in the context of the Vas'Y Filles! project. The forms had different items but measured the exact same content domains. In total, there were four combinations of forms and administrations: (a) form B with TL – EGMA B TL -, (b) form M with TL – EGMA M TL -, (c) form B without TL – EGMA B -, and (d) form M without TL – EGMA M (see Figure 2). Each girl received two scores (B and M), and these scores were used to answer the first and second research questions.

Figure 2. Administration of the EGMA and EGMA TL



To answer the third research question I took notes of all the translingual administrations, and tape-recorded 40 of them. These 40 administrations were randomly chosen, with three of them turning inaudible and not being included in the final analysis. The objectives of collecting these data were to obtain relevant information about the context in which these assessments took place, and to document the ways in which students used language during the administration. In addition, I conducted brief semi-structured interviews to the eight teachers in their respective schools. The interviews were conducted in French tape-recorded, and took no longer than five minutes each. The objectives of the interviews were (a) to broadly understand teachers' views on how to evaluate their students' math knowledge, and (b) to gain insight on the degree of alignment between the

EGMA TL and both, their views and practices in the mathematics classroom. I was interested in gaining a better understanding of the evaluation practices that teachers engaged in the classroom. I was also interested in gaining a better understanding of the interplay between the use of languages in the mathematics classroom. The questions that guided these interviews are shown in Table 6.

Next, I conducted a brief semi-structured interview to the enumerators that worked with me collecting the data. I was interested in getting a general opinion about the translingual administration of the EGMA and I was not looking to dig deeper into the subject. In particular, the objective of the interviews was to gain insight on the practical benefits or difficulties that the translingual administration may bring. To simplify the theoretical framework, I labeled and referred to the current EGMA administration as the “traditional administration” and to the translingual EGMA as the “new administration”. The questions that guided these interviews of enumerators are listed in Table 7.

Table 6. Original Set of Questions for Teachers

Number	Question
1	Comment est-ce que tu sais que tes élèves ont appris des connaissances de mathématiques ? [How do you know that your students learnt math content ?]
2	Quelle-est, a ton avis, la meilleure façon d’évaluer la connaissance de mathématiques de tes élèves ? [In your opinion, which is the best way to evaluate the level of math knowledge of your students?]
3	Comment est-ce que tu utilises les résultats des évaluations et/ou devoirs pour modifier la planification des ton cours de mathématiques ? [How do you use the results of tests and/or homework to plan your math course?]
4	Ton cours de mathématiques est donné dans quelle langue ? [In what language is your math class taught?]

Table 7. Original Set of Questions for Enumerators

Number	Question
1	Quelle façon d'administrer l'EGMA – traditionnelle ou nouvelle – est-ce que tu préfères ? Pour quoi? (Which EGMA administration – traditional or new – do you prefer? Why?).
2	A ton avis, quelle est la principale différence entre les deux administrations ? (In your opinion, what is the main difference between both types of administration ?)
3	Est-ce que t'es senti à l'aise en administrant l'EGMA à la façon nouvelle ? Pour quoi ou pour quoi pas? (Did you feel comfortable administrating the EGMA with translanguaging? Why or why not?)

3.3 Data Analysis

3.3.1 First Research Question

To answer the first research question, I began by analyzing the distributions of scores. I conducted visual analyses to look at the distributional shapes and used Kolmogorov-Smirnoff tests to evaluate the extent to which the distributions met normality. Then, I used both Kolmogorov-Smirnoff and Anderson-Darling tests to examine the differences between the distributions of the EGMA and EGMA TL administrations. This analysis was performed at the task level and at the total score level. After analyzing the characteristics and differences between the EGMA and EGMA TL scores, I conducted analyses of variance (ANOVAs).

The dependent variable in all of the ANOVAs was the average total EGMA score (average of baseline and midline total scores). The first model (see equation 1) had two fixed factors: (a) the experimental condition and (b) the grade level. Both these factors were treated as fixed as I was not attempting to generalize beyond the observed grade levels. In terms of results, I was interested in looking at the significance of the experimental condition and of the interaction between the experimental condition and grade level. To

that end, I planned to conduct two complex tetrad comparisons to see if the translingual administration made a difference for lower grades and/or higher grades (see equation 2), using the Holm's method to control for the family-wise Type I error rate. If significant, I would follow up with simple comparisons.

Equation 1. Two-Factor Fixed ANOVA Model

$$Y_{ijk} = \mu + \alpha_j + \beta_k + (\alpha\beta)_{jk} + \varepsilon_{ijk}$$

- Y_{ijk} is the average EGMA score for person i in experimental group j (0=EGMA, 1=EMGA TL), and grade level k (3rd, 4th, 5th or 6th)
- μ is the grand mean
- α_j is the main effect of taking the test with translingual administration
- β_k is the main effect of being in grade level k
- $(\alpha\beta)_{jk}$ is the interaction effect for the combination of experimental condition and grade level
- ε_{ijk} is the random residual error for individual i in groups j and k

Equation 2. Planned Multiple Comparisons for the Two-way ANOVA

$$\psi_1 = \frac{\bar{Y}_{.03} + \bar{Y}_{.04}}{2} - \frac{\bar{Y}_{.13} + \bar{Y}_{.14}}{2}$$

$$\psi_2 = \frac{\bar{Y}_{.05} + \bar{Y}_{.06}}{2} - \frac{\bar{Y}_{.15} + \bar{Y}_{.16}}{2}$$

These contrasts are compared to a t value of $t_{\alpha/2, (N-J*K)}$, where the standard error for each contrast is provided by:

$$s_{\psi_h} = \sqrt{MS_{within} \left(\sum_{j=1}^J \sum_{k=1}^K \frac{c_{hjk}^2}{n_{jk}} \right)}$$

And the effect size for each contrast was given by Cohen's d:

$$d_h = \frac{\Psi_h}{\sqrt{MS_{within}}}$$

Where:

- Ψ_1, Ψ_2 are the planned contrasts
- \bar{Y}_{jk} are the means for individuals in group j and grade k
- c_{hjk}^2 is the square of the coefficient for mean jk in the contrast h
- n_{jk} is the number of individuals in group jk

The second model (see equation 3) had three factors: (a) experimental condition, (b) grade level, and (c) home language. Under a translangual orientation to languages, we should conceive this last variable (home language) as random or more precisely, as continuous. While theoretically sound, this option was not feasible for this analysis. With regards to the first option (language as a random variable), the sampling procedure was such that home language could not be treated as a random variable. The sampling was not random at a home language level, and therefore, I could only treat this variable as fixed and generalize the associated results to the particular socio-lingual context in which this study was conducted. With regards to the second option (language as a continuous variable), the way in which we measured home language did not build in continuity. Therefore, I could not treat this variable as continuous and instead, treated home language

as a fixed factor, with two levels corresponding to the local language (Lingala) and to “other languages” (namely French and Mixed/Bilingual/Both). The reason for this latter grouping is that only one girl reported using French, and the rest of the girls reported speaking either Lingala or both French and Lingala at home. In terms of results, I was mostly interested in looking at the significance of the experimental condition and of the interaction between the experimental condition and home language. To that end, I planned to follow up with simple comparisons between the EGMA TL and EGMA groups for students who spoke different languages at home. To control for the family-wise type I error, I also planned to use Holm’s.

Equation 3. Three-Factor Fixed ANOVA Model

$$Y_{ijkl} = \mu + \alpha_j + \beta_k + \gamma_l + (\alpha\beta)_{jk} + (\alpha\gamma)_{jl} + (\beta\gamma)_{kl} + (\alpha\beta\gamma)_{jkl} + \varepsilon_{ijkl}$$

- Y_{ijk} is the average EGMA score for person i in experimental group j (0=EGMA, 1=EMGA TL), grade level k (3rd, 4th, 5th or 6th), and language group l (L,O)
- μ is the grand mean
- α_j is the main effect of taking the test with trasnlingual administration
- β_k is the main effect of being in grade level k
- γ_l is the main effect of speaking language l at home
- $(\alpha\beta)_{jk}$ is the interaction effect for the combination of experimental condition and grade level
- $(\alpha\gamma)_{jl}$ is the interaction effect for the combination of experimental condition and home language

- $(\beta\gamma)_{kl}$ is the interaction effect for the combination of grade level and home language
- $(\alpha\beta\gamma)_{jkl}$ is the interaction effect for the combination of experimental condition, grade level, and home language
- ε_{ijkl} is the random residual error for individual i in groups j k and l

For the two-way and the three-way ANOVA, I used type III sum of squares to evaluate the effect of each main effect and interactions. I chose this approach since the number of individuals in each group varied. While a type II sum of squares approach would have been more powerful in the absence of interactions, I had no prior assumption about the existence of interaction, and went with a more conservative approach.

Last, I conducted a multiple linear regression analysis using two factors and two covariates. As factors I used the experimental condition and the language spoken at home. Similarly to the ANOVAs, the variable regarding experimental condition, distinguished students who took the EGMA from students who took the EGMA TL, while the variable regarding home language distinguished students who spoke Lingala at home, from students who spoke mixed varieties or other languages. In terms of covariates, I used grade level and age, as they potentially affected EGMA scores. In the regression model, grade level was treated as a continuous variable because the assumption of equal spacing held, and because it was not a variable of interest; as a control variable, grade level was better used continuously.

To test the effect of each predictor, I used a hierarchical approach to regression. I set up a full model with all the relevant predictors. Then, I removed each predictor and saw

the difference in terms of multiple correlation (R^2), to evaluate its effect on the equation. I tested each coefficient against the mean sum of squares of the full model, to be conservative. This hierarchical approach, equivalent to the type III sum of squares in an ANOVA/ANCOVA design, did not use the same coding as the ANOVAs. Instead, for this analysis and for the final regression analysis, I used a different code system (see Table 8). The reason for this was to obtain coefficients that were more meaningful in light of the questions I was trying to answer. The first factor was turned into a dummy variable: 0 for the traditional administration, and 1 for the TL administration. The home language factor was coded as an unweighted effect, mainly because I was interested in the difference between the groups (Other – Lingala) rather than in the individual effect of any of the groups. The interaction term in the final equation (see equation 4) reflects this coding system.

Equation 4. Final Linear Regression Model

$$Y_i = \beta_0 + \beta_1 * TL + \beta_2 * C_{HL} + \beta_3 * TL * C_{HL} + \beta_4 * Age_c + \beta_5 * Grade_c + \varepsilon_i$$

Where:

- TL is the dummy variable for the experimental condition
- C_{HL} is the unweighted effect for the home language variable (Other – Lingala)
- $TL * C_{HL}$ is the interaction term between the factors
- Age_c is the centered age
- $Grade_c$ is the centered grade level
- ε_i is the residual term

Table 8. Coding Schemes Used in the Regression Analysis

	Final Regression		
	TL	C_{HL}	$TL * C_{HL}$
Lingala EGMA	0	-1	0
Lingala EGMA TL	1	-1	-1
Other EGMA	0	1	0
Other EGMA TL	1	1	1

3.3.2 Second Research Question

To answer the second research question, I estimated the reliability coefficient for the translingual administration and the traditional administration. For each type of administration, reliability was estimated as the correlation between the aggregate scores from forms B and M. In this context, this estimation was more appropriate than using internal consistency estimates (e.g. Cronbach alpha) for a couple of reasons. First, the EGMA has a bizarre use of time, where some yet not all of the tasks are timed. Time constraints may add dimensionality to the measure, and dimensionality may artificially inflate internal consistency reliability, as score consistency depends on both, the correlation between item scores and the dimensionality of the assessment (Cortina, 1993). Second, timed tasks may interact with students' proficiency in a way that diminishes internal consistency reliability estimates: if the students affected by timing rules are all of a given proficiency, the reliability of the whole instrument for the whole set of examinees, will be underestimated using internal consistency measures. Third, while the EGMA is not a multiple-choice assessment, there could be some guessing associated with its time rules, affecting internal consistency measures even further (Attali, 2004).

The appropriate reliability estimate depends on the particular error-producing factors in a given situation. An alternative to parallel forms reliability would have been test-retest estimates, but this was not ideal since it was logistically more cumbersome and did not add value. The most appropriate estimates for this particular context, were parallel forms estimates. Therefore, I estimated reliability for both the EGMA and the EGMA TL using parallel forms methods, where each estimate corresponds to the correlation between two forms B and M (see equation 5). I computed reliability estimates for the whole tests as well as for each task. Important to mention, order and learning effects were not relevant since I alternated the forms across the students. Thus, it was unlikely to observe systematic error due to the order in which the forms were administered, protecting the accuracy of the estimates.

Equation 5. Parallel Forms Reliability

$$\rho_{XX'} = \rho(X, X') = \frac{\sigma_T^2}{\sigma_X^2}$$

Where:

- X= Form X
- X'= Form X'

3.3.3 Third Research Question

To answer the third research question, I mostly used qualitative analyses. First, I conducted an interpretation analysis of the field notes that I took as I administered the translingual EGMAs. The purpose of the analysis was to recover essential characteristics

of the administration. My field notes were mixed with planning notes, with personal notes, and varied widely from day to day. While the main objective of analyzing my field notes was to recover contextual details, reproducing these notes in this study, depicts my own translanguaging and serves to highlight how a focus on meaning rather than on form is critical to make sense of content of bilinguals, such as myself. For this reason, I tried to reproduce the notes as faithful as possible. The nature of this analysis was interpretive.

Second, I conducted a structural analysis of the students' test-speech, based on the tape-recordings. By test-speech I mean the speech of girls as they took the assessment, which is not representative of these girls' speech, but rather treated as a very specific instance of it. The objective of this analysis was to uncover the ways in which girls used language during the EGMA TL administration, to discover potential patterns of test-speech across the girls. I exclusively focused on the code-switching aspect of speech, but further analyses of these recordings may uncover other characteristics that are relevant to the implementation of translanguaging in assessments. I explored two main themes: (a) the extent to which girls engaged in flexible bilingualism, and (b) the characteristics of girls' flexible bilingualism. The nature of this analysis was descriptive.

To answer the third research question I also analyzed the eight interviews to teachers and the interview to one of the enumerators. With regards to teachers' interviews, the analysis consisted on identifying a set of themes that could shed light on the beliefs that teachers held with regards to assessing mathematics' knowledge of students. In addition, the analysis of the interviews allowed me to gather more information about the reality of the classrooms, information that was also considered in the evaluation of the appropriateness of the translingual administration of the EGMA. The nature of this analysis

was interpretive. Last, with regards to the enumerator interview, the analysis consisted on judging the feasibility of translingual assessments in this particular context, as evidenced by the experience and opinions of the enumerator.

Table 9. Summary of Methods by Research Question

No	Research Question	Type of Method	Method of Inquiry
1.1	Does a translingual administration of the EGMA assessment have an impact on math achievement of Congolese girls, as evidenced by a significant difference in their scores when compared to the traditional administration?	Quantitative	Kolmogorov-Smirnov test, Anderson-Darling test, ANOVAs, and Linear Hierarchical Regression Analysis
1.2	Does a translingual administration of the EGMA assessment allow us to obtain more reliable information about Congolese girls' math knowledge, as evidenced by higher reliability estimates when compared to the those obtained from the traditional administration?	Quantitative	Parallel Forms Reliability Estimation of Wordy Tasks, Non-Wordy Tasks, and EGMAs.
2.1	Is the translingual administration of the EGMA assessment appropriate for the context in which this data was collected, as evidenced by: <ul style="list-style-type: none"> • The characteristics of the context in which the assessments were administered • The level and ways in which girls engaged in flexible bilingualism • The alignment with teachers' practices beliefs around how to assess math proficiency • The feedback from enumerators who administered the EGMA 	Mixed	Interpretive analysis of field notes, structural analysis of tape-recordings, interpretive analysis of teachers' and enumerator's interviews.

CHAPTER 4

RESULTS

4.1 Does a translingual administration of the EGMA assessment have an impact on math achievement of Congolese girls, as evidenced by a significant difference in their scores when compared to the traditional administration?

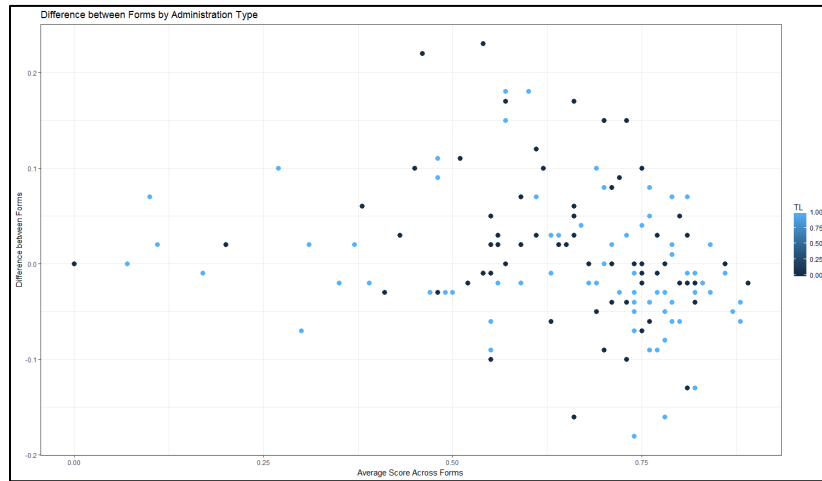
To answer this question, I computed the total scores per task and form. The EGMA used in this study had 5 tasks. The first corresponded to *number identification*, where students had to name as much numbers correctly as they could in 60 seconds. Students were exposed to at least 5 numbers, up to a maximum of 20 numbers. The scoring of this task could be of two types: correct numbers per minute (CNPM) or percent correct. I computed both scores, but only the percent correct was used in the computation of a total EGMA score. The percent correct for this task – as well as all other timed tasks – was computed over the number of items attempted. In turn, the second task corresponded to *number comparison*, where students had to identify the largest number among two. This task had 10 items, and was untimed; therefore, only the percent correct score was computed. Next, the third task corresponded to *missing numbers*, where students had to name the missing number in a pattern of four numbers. This task was also untimed, so I computed the percent correct, estimated over the total number of items (10). The fourth task corresponded to *additions*, where students had to add up to 20 combinations of numbers, depending on their performance and speed. Because this item was timed, I computed two scores: correct additions per minute (CAPM), and the percent correct, over number of items attempted. The fifth and last task corresponded to *subtractions*, where students had to subtract up to 20 combinations. The task was timed and therefore, I

computed two scores: the fluency score – correct subtractions per minute (CSPM), and the percent correct, over the number of items attempted. Lastly, to compute an overall score per form, I weighted each task equally, and computed an average percent correct score across the five tasks. For more details on each of the forms, see Appendices A and B.

To analyze the data, we took the average score between the baseline and midline EGMA forms. We could have equated these two forms, but the difference between the scores was small and evenly distributed across the level of performance and type of administration. In Figure 2, we plotted the difference between the baseline and midline scores against the average score, distinguishing those students who took the EGMA (black points) from those students who took the EGMA TL (light blue points). As observed in the figure, the differences do not seem to follow any functional pattern with regards to the performance level or the type of administration (translingual or traditional). Because we are not interested in comparing baseline to midline performance, but rather EGMA to EGMA TL, we do not need to equate forms and we can take an average score between them, as they are not introducing any systematic bias to the analysis.

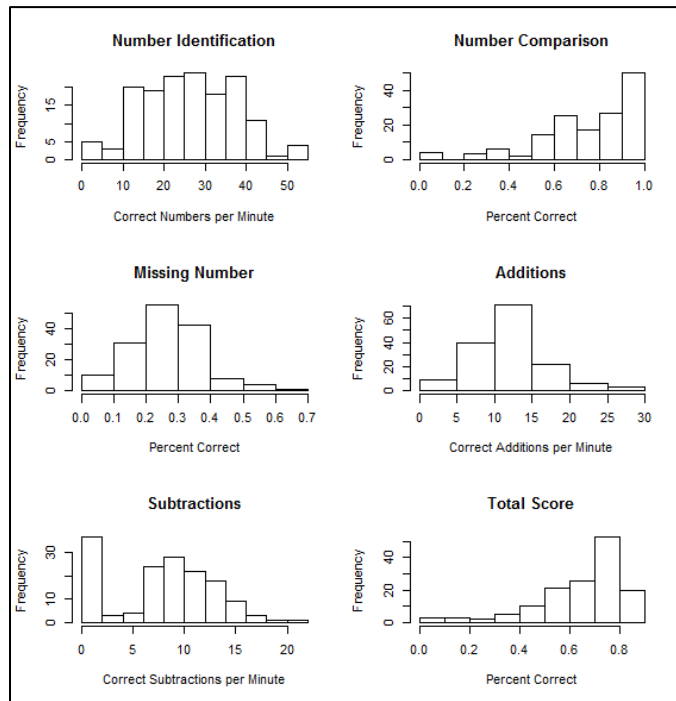
Taking the average scores between baseline and midline forms, we looked at the distributional properties of the tasks and totals. Figure 3. displays the histograms by tasks and totals for all the data; the histograms include information on both the EGMA and the EGMA TL. As observed, there seems to be a wide variety with regards to the distributional characteristics, across tasks. While three distributions show some signs of normality (number identification, missing number, and additions), the distribution for number comparisons is skewed to the left, the distribution for subtractions is slightly bi-modal (one

Figure 3. Differences in Scores between Forms by Administration Type



bin of zero scores plus a rightly skewed distribution), and the distribution for total scores is skewed to the left. The distributions of the EGMA tasks, across both forms and types of administration, do not look normal.

Figure 4. Distribution of Scores by Tasks and Totals



The results do change slightly if we take into consideration the type of administration; that is, if we consider whether students took the EGMA or the EGMA TL. Using a Kolmogorov-Smirnov test (K-S test), we tested normality for each task and for the total scores, distinguishing the two types of administration. The K-S test is appropriate for continuous data, and requires relatively small sample sizes (Press, Teukolsky, Vetterling, & Flannery, 1992). Table 10. displays the results of the K-S tests, where those results that indicate non-normal distributions are highlighted in bold (p values less than 5%). As noted, the number identification task, the missing number task, and the addition task look normal. The number comparison task and the subtraction task look non-normal, for the EGMA administration. The total EGMA score distribution looks normal for the traditional EGMA administration.

Table 10. Test of Normality for Tasks and Total Scores by Type of Administration

Task	Type of Score	Normality of EGMA (P-Value)	Normality of EGMA TL (P-Value)
Number identification	Correct numbers per minute	0.42	0.37
Number comparison	Percent correct	0.00*	0.16
Missing number	Percent correct	0.09	0.05
Addition	Correct additions per minute	0.37	0.23
Subtraction	Correct subtractions per minute	0.02*	0.16
Total	Percent correct	0.33	0.01*

* p<0.05

While normality is a desirable characteristic, it is not what I was mostly interested in detecting. The first research question had to do with understanding to what extent the translingual administration changed the properties of the scores. Therefore, I looked at the distributional differences between the EGMA and the EGMA TL. To do this, I also used the K-S test given the size and characteristics of the sample. Technically speaking, the K-S test is based on the value of a statistic (D) that measures the largest absolute difference

between two cumulative distribution functions. It is a powerful test in the sense that it does not require large sample sizes ($n \geq 4$), but it is sensitive to outliers and tends to be more sensitive around the median value than at the extremes of the distributions (Press et al., 1992). Therefore, to increase the power in the tails, I also compared the EGMA and EMGA TL distributions using a test called the Anderson-Darling test (A-D test). Table 11. displays the results for both the K-S comparisons and A-D comparisons: as observed, no test had a p-value < 0.05 . This means that in no case we could reject the null hypothesis that different pairs of scores (EGMA, EGMA TL) came from a same distribution. There is no evidence that the translingual administration changed the distributions of scores.

Table 11. Kolmogorov-Smirnov and Anderson-Darling Tests by Type of Administration

Task	Score	K-S Test (P-Value)	A-D Test (P-Value)
Number identification	Correct numbers per minute	0.19	0.22
Number comparison	Percent correct	0.76	0.47
Missing number	Percent correct	0.98	0.98
Addition	Correct additions per minute	0.89	0.60
Subtraction	Correct subtractions per minute	0.80	0.73
Total	Percent correct	0.09	0.10

The distribution of scores did not change at an aggregated level, but it is possible that the translingual administration had other effects on the scores for groups or cases. To understand the differences that the EGMA and EGMA TL at greater detail, I used ANOVAs. The first ANOVA included two fixed effects: treatment (EGMA or EGMA TL) and grade level (3rd grade through 6th grade). As shown in Table 12, the test showed that only grade level had a significant relationship with the average EGMA score. Because the F-test was not significant for the interaction between treatment and grade level, I did not follow up with the multiple comparisons, as intended.

After adding the home language variable, the results change slightly (see Table 13 for a full display of the results). Grade level remains a significant factor and treatment remains non-significant. Home language per se is not a significant factor either. And there is no significant interaction between grade level and home language. Yet, we do observe a significant interaction between the experimental condition and the home language: the effect of the translingual administration is contingent to the language spoken at home. This finding suggests that incorporating linguistic flexibility in the administration of mathematics assessments may have an impact on students' performance, depending their linguistic practices.

Table 12. Results of the 2-Way ANOVA

	DF	Sum of Squares	Mean Square	F	Pr(>F)
Treatment (TL)	0.02	0.02	0.767	0.767	0.38
Grade	1.27	0.42	18.62	18.62	0.00***
Treatment (TL): Grade	0.04	0.01	0.57	0.57	0.64
Residuals	135	3.08	0.02		

*** p<0.001

Table 13. Results of the 3-Way ANOVA

	DF	Sum of Squares	Mean Square	F Test	Pr(>F)
Treatment (TL)	1	0.00	0.00	0.02	0.88
Grade	3	1.19	0.40	17.68	0.00***
Home language (HL)	1	0.02	0.02	1.03	0.31
Treatment (TL): Grade	3	0.03	0.01	0.44	0.72
Treatment (TL): Home language (HL)	1	0.12	0.12	5.15	0.024*
Grade: Home language (HL)	1	0.00	0.00	0.21	0.64
Treatment (TL): Grade: Home language (HL)	0				

* p <0.05; *** p<0.001

Table 14 shows the simple comparisons between the EGMA and EGMA TL, for different language groups. We note that the translingual administration only affects the scores of the students who speak “other languages” at home (French or a mix between French and Lingala). The difference is such that the translingual administration increases

the mean in the total EGMA score by 0.4 percentage points, which is considerable. However, the small sample size for this group of students (n=9) prevents us from generalizing the results, at this point. The impact of the translingual administration did not have any significant effect on the Lingala speaking students.

Table 14. Simple Comparisons 3-Way ANOVA

	Difference in Means	Standard Error	Contrast	Pr(> t)
EGMA TL vs. EGMA - Lingala group	-0.02	0.03	-0.64	0.52
EGMA TL vs EGMA - Other languages group	0.39	0.16	2.48	0.01*

* p <0.05

To finish answering the question of whether the translingual administration had any effect on the scores of the EGMA assessment, was provided by means of a hierarchical linear regression. We first looked at whether the experimental condition (TL term), the home language factor (CHL term - contrast), and their interaction was significant, using a hierarchical approach. Table 15 displays the summary of these results. First, we observe that the TL term contributed significantly to the explained variance, given this new parametrization. This means that the experimental condition influenced the mean EGMA scores, after controlling for grade level and age, with this way of setting the variables. Second, we observe that the HL term did not have a unique contribution on the scores' variance, but that the interaction term between TL and HL was again, significant. This means that the translingual administration had a different effect depending on the language which students spoke at home, after controlling for age and grade level. This is the most important outcome, consistent across different parametrizations. To look at the nature and magnitude of these effects, we need to look the results for the full model.

Table 15. Unique Contribution of the TL and the HL Terms on the Regression Analysis

	Unique Contribution	Degrees of Freedom	F Test	P-Value
Treatment (TL) term	0.11	1.00	4.86	0.03*
Home language (CHL) term	0.05	1.00	2.22	0.14
Interaction term (TL:CHL)	0.11	1.00	4.89	0.03*

* p <0.05

The final regression model included two covariates, the two factors, and their interaction. The R^2 for this regression was 0.29, which is moderate in the context of social sciences. The main results for the final regression are provided in Table 16: coefficients, standard error of the coefficients, t statistics, and the p-value for each of the p statistics. As observed, the results for the TL, HL, and interaction terms are consistent with those displayed in table 15. In addition, we observe that the regression term for the intercept and for the grade level, are significantly different than zero. The age covariate did not add any explanatory power to the model.

Table 16. Results for the Final Multiple Regression Model

	Coefficient (B)	Standard Error of the Coefficient	T Value	Pr(> t)
Intercept	0.53	0.08	6.91	0.00***
Treatment (TL)	0.18	0.08	2.20	0.03*
Home Language Term (CHL)	-0.11	0.08	-1.49	0.14
Age (centered)	0.00	0.01	-0.35	0.73
Grade (centered)	0.08	0.02	5.14	0.00***
Interaction Term (TL:CHL)	0.18	0.08	2.21	0.03*

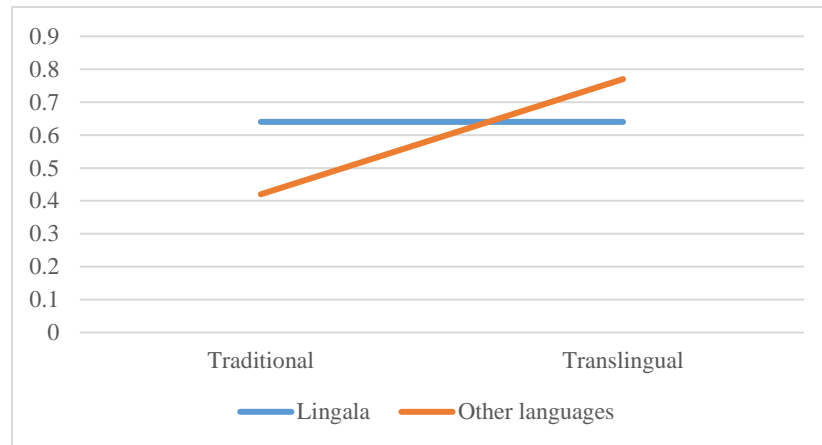
* p <0.05; *** p<0.001

Finally, table 17 shows the adjusted means for different groups. Table 17 shows that the EGMA TL resulted in higher means for the “Other Languages” group and in no significant effect for the Lingala group. However, the sample size of the “Other Languages” group was too small and not randomly sample, so these results, despite being encouraging, cannot be generalized. A more meaningful display of the results is presented in Figure 5.

Table 17. Adjusted Means

Adjusted Means	HL= Lingala	HL=Other
TL	0.64	0.77
Non TL	0.64	0.42

Figure 5. Adjusted Means – Graph



4.2 Does a translingual administration of the EGMA assessment allow us to obtain more reliable information about Congolese girls’ math knowledge, as evidenced by higher reliability estimates when compared to the those obtained from the traditional administration?

To answer this question, I computed parallel forms reliability analyses (correlations) for the scores of each task and for the whole test, separately for the two different forms. Table 18 displays these results: the second column in Table 18 provides the reliability estimates for the translingual administration, while the third column provides the reliability estimates for the traditional administration. We observe a couple of things from Table 18. First, every single task has a higher reliability estimate when using the translingual administration as the basis for the calculations. Second, the differences are high for every task except for additions and subtractions, which present low differences in

the estimate across administrations (0.1 and 0.3 points respectively). Third, most of the tasks, except for the missing number task, present relatively high estimates across both administrations.

Table 18. Parallel Forms Reliability Estimates

Task	EGMA TL	EGMA
Number identification	0.91	0.77
Number comparison	0.80	0.68
Missing number	0.58	0.44
Additions	0.74	0.73
Subtractions	0.90	0.87
Total	0.95	0.89

To make more sense of these results, we can look at the percentage of variance explained by each form on the other (e.g. baseline on midline) by squaring the reliability estimate. Table 19 presents these results. As observed, there is a clear increment in the proportion of shared variance for the number identification task when switching to the translingual administration (increment goes from 59% to 83%). The increment is less pronounced yet noticeable for the number comparison, missing number, and subtractions task, as well as for the total scores. The only task that shows almost no variation with regards to the proportion of shared variance is additions.

Table 19. Proportion of Shared Variance between Forms

Task	EGMA TL	EGMA
Number identification	83%	59%
Number comparison	64%	46%
Missing number	34%	19%
Additions	55%	53%
Subtractions	81%	76%
All items	90%	80%

Now, the distribution of correlations is non-normal; it is a negatively skewed distribution that depends on the sample size and true population correlation. Therefore, it is difficult to judge whether the differences in correlations (or in proportion of variance explained) were significant or not, without a transformation. To test for statistical significance of the differences in reliability estimates, I used a z-transformation. Table 20 shows the results of the transformations, as well as the results of the statistical test to compare the differences (a z-test). As observed, only the difference in correlations for the number identification task and for the total test, were statistically significant. The first result is not surprising, given the large increment observed in Table 20. The second result is slightly less intuitive, as the difference in reliability for the total score was not larger than for other tasks. However, the difference is more meaningful since the total score includes a lot more items than a single task; the result makes absolute sense.

Table 20. Z-test for Differences in Correlations

	Z Fischer			
	EGMA TL	EGMA	Z for difference	Pr (> Z)
Number identification	1.53	1.02	3.04	0.00
Number comparison	1.10	0.83	1.60	0.11
Missing number	0.66	0.47	1.14	0.21
Additions	0.95	0.93	0.13	0.40
Subtractions	1.47	1.33	0.83	0.28
All items	1.80	1.43	2.19	0.04

4.3 Is the translingual administration EGMA assessment appropriate for the context in which this data was collected?

4.3.1 The Context

The EGMA assessments took place in the urban area of Mbandaka, but the label of *urban* does not translate into the concept of *urban* imagined from developed countries. Mbandaka has some paved streets but most of the streets are not paved, and its downtown is comprised of few streets with small businesses: while remarkably crowded and dense, and therefore technically a city, Mbandaka shares many features characteristics of rural regions in other parts of the world. The schools that we visited were in dispersed areas of Mbandaka, miles away from downtown, in rather isolated communities. The weather conditions were extremely hot, dusty, and with unexpected and dramatic rains. It is important to grasp that schooling conditions in this area are very difficult, as students must deal with harsh weather conditions, on top of the challenges and injustices that poverty entails.

The EGMA TL was administered by two enumerators – myself, and one of two local staff. We typically arrived at the schools early in the morning, and coordinated with teachers to select the girls to be assessed. At each school, we were provided empty classrooms with chairs and desks, so that we could administer the assessment in a quiet and proper space. However, my presence in the schools was distracting, as little girls were not used to see a “mumdele” or a “white woman” (I am Hispanic) around. The distraction and excitement was such that many girls left their classrooms and looked through the window as we were administering the assessment to other students. Many of them could not hold their excitement and interrupted the assessment several times, mostly to say

“mumdele mumdele” and run away. My presence was not unnoticed and the quietness of the administration was consequently challenged.

The instructions of the EGMA were printed in Lingala and French to ensure some degree of standardization in the delivery of instructions. However, the most salient theme across my field notes relates to the inherent difficulty that the context poses to any standardized assessment. One of the things that I noted repeatedly was that we could not effectively deliver instructions in a standardized way. To be clear, the translingual administration did not aim for standardization, so variety across administrations was not unwanted, but as we administered more and more assessments, I simply could not imagine a standardized administration for the EGMA or any other educational assessment, in this context. Some of my most informative notes on this regard are:

1. *“None of the instructions was repeated exactly in the same way and it feels more natural and appropriate for this context”;*
2. *“Interruptions are so normal, they [enumerators] do not even realize the effect that this has on standardized assessments”;*
3. *“This administration is as standardized as possible”;*
4. *“Imposible estandarizar instrucciones”* [**English:** It is impossible to standardize instructions];
5. *“Standardization is def. [definitely] not useful in this context”.*

One of the core characteristics of this context was the variety with which the EGMA instructions were delivered across administrations. This lack of standardization was common to other pieces of the administration. For example, I noted that the way in which enumerators used the chronometers and timed the girls’ responses was highly irregular. Some of the notes that highlight this issue include the following:

6. “The enumerator forgets easily [easily forgets] that she has to time the exercise”;

7. “It is worth to be flexible w/th. 5 secon.[with the 5 seconds rule]”;

8. “Sonó un celular -> estos retrasos de tiempo {o por ejemplo que las niñas afuera estén molestando} no se incluyen en el tiempo de administración. Confiar en la estandarización de este instrumento es difícil. Me parece que la administración oral y local lo hace imposible” [**English**: A cell pone rang -> these delays {or for example, that girls are outside distracting} are not included in the time of administration. To trust the standardization of this instrument is difficult. It seems to me that the oral and local administration makes this impossible];

9. “I do not stop them if they are en train de donner les responses” [**English only**: I do not stop them if they are in the process of giving the answers].

It seemed to me that timing tasks did not make a lot of sense in this context. The reason is that there were many barriers to properly accounting for time on tasks, barriers that felt unsurmountable. The irregularity that characterized the deliverance of instructions and the timing of the administrations, seemed to index a larger cultural phenomenon: that irregularity could be an essential characteristic of the Mbandaka reality. I did not count the number of interruptions that we had, or did not make an inventory of the types of the disruptions I witnessed while administering the EGMA, but we had plenty of pauses due to unexpected events. We also had to change plans many occasions, for similar reasons. At one point I wrote 10. “Tuvimos OTRA interrupción. On va s’arreter pendant la recreation” [**English**: We had ANOTHER interruption. We will stop during recess]. This sentence shows my frustration with the whole issue of interruptions, and the need of changing plans for that particular day; it was very difficult to stick to any pre-determined plan for any given day or administration.

My frustration was something that I did not understand at the time of administering the EGMAs. Looking backward, it was difficult for me to understand why it was so hard to achieve regularity in the administration of the assessments. I tend to associate

standardization with discipline and focus, but I had never reflected on the issue of why standardization is so “natural” to some of us. Standardization and regularity are most likely the outcomes, not the precursors, of formal schooling. And standardization and regularity may not always be the product of formal schooling, since in contexts like Mbandaka, variation is the norm, and this variation affects schooling. For example, teachers were often sick and despite its repeated occurrence, there was no clear rule as to what to do whenever this happened. As I wrote: 11. *“There are always lots of absences. [Today] There is a sick teacher and so students go home? [went home] Instead of sharing the [a] classroom”*. This note describes that on a given occasion, students were sent home because their teacher was sick, which affected our planning for that day. Standardization and regularity cannot be taken for granted, and may be rather rare in contexts that are governed by more basic priorities than education, such as nutrition and health.

This lack of standardization was reflected more subtly in other behaviors. For example, it was difficult for the enumerators to remember to turn off the cellphone while administering the assessment, which often rang. It was absolutely not relevant for them whether the phone rang or not, aside from being an explicit requirement of the EGMA administration. This relative unconcern towards the regularity of the administration shows that enumerators had not fully internalized the concept of standardization or its relevance for standardized assessment. On the other hand, with regards to the girls, while all of them were administered the assessment twice, it was rare for them to recall the instructions during the second EGMA. We had to repeat the instructions on every single occasion, which felt strange given the proximity between both administrations. I had many notes regarding this issue, the most important being:

12. *“Girls are as surprised with instructions as the first time...”*

13. *“We repeat the instructions every time and it sounds strange to me, to say the exact same thing”*

14. *“I tried to go without saying the instructions but she didn’t remember. From 10 mins ago!”*

15. *“She didn’t understand, so we had to repeat instructions & add explanation ----- This happens a lot.”*

Another relevant characteristic that I noted during the administration of the EGMA TL, was that most of the girls were shy. While this is not the focus of this study, we should ask whether these assessments make sense in a context where girls are afraid to individually speak out. It is common practice in DRC to use “*comptines*” (short songs with rhymes) and group-based activities to engage and monitor the performance of students in the classroom. An individual oral assessment may be too novel of an experience for these timid girls, and my experience suggests that the question of whether individual assessments are appropriate for this context at all, is worth exploring.

4.3.2 Flexible Bilingualism Among the Assessed Girls

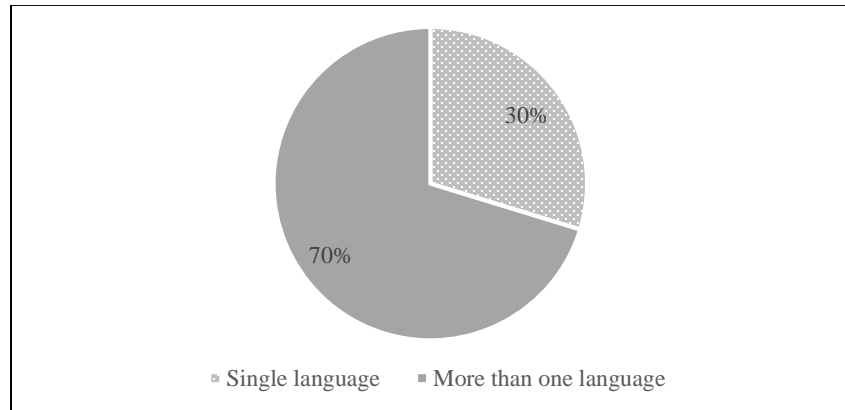
The EGMA assessment was not ideal to explore girls’ speech, as the required answers were not long enough to get stable notions of their speech structure. Some of the tasks could be answered with single numbers. There were only two tasks in which girls could use and typically used sentences to provide their answers: number comparison and missing number. In the administration of these tasks, we encouraged girls to respond using sentences rather than single words (numbers). For the number comparison task, we encouraged girls to respond using sentences such as “Number A is greater than Number B”. For the missing number task, we encouraged girls to say “The missing number is X”. However, even if girls used sentences to provide their answers, these sentences were

incredibly bounded and therefore, we could not conduct a speech analysis but rather identify general trends within the use of language when communicating math knowledge.

Based on the test-speech of 37 girls, we can learn something about the ways in which these students translanguage. To analyze the recordings, I used a single main linguistic variable of interest, namely, whether girls used lexical items from French, from Lingala, or from both languages. The analysis of the speech was mostly quantitative, in the sense that I classified the excerpts according to their lexical composition: French, Lingala, Mixed. However, I also provide some concrete examples of the ways in which students used language to answer some of the questions, to illustrate potentially stable translingual behaviors among this population of students. These examples will be discussed at the end of this sub-section.

The analysis of the words used by students was done at three different levels: at a test level, considering the administration of both forms B and M; at the form level, considering the administration of each form separately; and at the task level, using the information from both administrations. With regards to the first level, the main issue was to find out whether girls used words from both languages to provide their responses, or used words from one single language. Figure 4 shows that 70% of these girls used words from more than one language to provide their responses. While I acknowledge the monoglossic nature of this analysis – looking at languages as separate and distinguishable entities –it does shed light on how inappropriate it is to ask children to express themselves in one single language. Even in a short and low-speech assessment like the EGMA, most children deploy bilingualism when answering the items.

Figure 6. Number of Languages Used in Providing Answers



Within those girls who used a single language to provide their responses, across both forms³, all of them used French only. This is relevant as there could be a causal link for such behavior: French is the official language of instruction from 3rd grade on but it takes a bit to have a French-only class. It could be that these girls (n=11) were all 5th or 6th graders, yet the way in which these girls were tape-recorded was random, to the extent that I lost any reference with regards to their age or to their grade level. It was not possible to verify whether these girls were from upper grades, but it is likely.

After analyzing the variable (lexical items) at the second level, I realized that there was variation across forms. A total of 25 girls (71%) deployed similar behaviors across both forms (baseline and midline). While most girls showed stable patterns of lexical use, 29% of the tape-recorded girls did not, which is a large percentage of girls. More importantly, when we look at the girls who showed variation across forms, we realize that there is a pattern in the directionality of the change. Only one of the girls used French only in the first administration and Mixed language in the second administration: 9 out of 10

³ Few girls responded to only one form. This analysis includes these girls.

girls used mixed language in the first form and French only in the second form. This pattern shows how translanguaging may have acted as a scaffolding device. It also aligns with the way in which bilingualism is enacted in the DRC: girls are supposed to learn in French, yet teachers can use mother tongues in earlier grades to ensure understanding. Therefore, these patterns (Mixed -> French) may have some relationship with the language policies operating in the Mbandaka schools.

More insightful results appear when we look at the third level of analysis: the task level. Table 21 shows the variation across forms by tasks. We observe that most of the differences took place when responding to the “wordy” tasks: number comparison and missing number. This relationship does not seem spurious but genuine: it is in longer excerpts that we can realize the flexible use of language that these girls deploy in their speech. However, it is important to know that a large percent of girls responded to the number comparison and missing number tasks, using numbers only: 77% and 54% of the answers to these tasks, were of this kind. This was consequential as it restricted the possibilities to observe translanguaging any further.

Table 21. Lexical Items across Forms by Tasks

	Same	Different
Number identification	34	1
Number comparison	26	9
Missing number	27	8
Addition	31	4
Subtraction	33	2

Table 22 shows more patterns between tasks. We observe that there are three tasks were students rarely or never gave responses using words from more than one language: number identification, additions, and subtractions. This is not surprising given that these

tasks could be answered using single words (indeed, numbers). Table 22 also shows that there were some students who used Lingala only in some of their answers. In particular, there were two responses provided in Lingala: one for number identification and one for additions. Responding in Lingala only did not take place at the test level, or at the form level, but it did take place at the task level. Therefore, we can state the using Lingala only was a rare occurrence, but we cannot efface Lingala from the universe of response possibilities.

Table 22. Lexical Items by Language and Task Across Administrations

	Number Identification	Number Comparison	Missing Number	Additions	Subtractions
Lingala	1.00	0.00	0.00	1.00	0.00
French	70.00	39.00	46.00	67.00	69.00
Mixed	0.00	33.00	25.00	3.00	2.00

The numbers do not add to 74 as not all students answered all tests/items

When we analyze the lexical items for the number comparison and missing number tasks, we observe that there is a stable pattern of language use: numbers are mostly enounced in French, and operations or comparative language are mostly enounced in Lingala. Almost every girl who used mixed language to respond to these items followed that pattern. Some examples of responses that girls gave to these tasks are presented in Excerpts 1 and 2. It is unclear to what degree the translingual administration influenced this outcome. To be clear, there was no statistical relationship between the language of the instructions and the language of the responses. Yet some tape-recordings show some relationship between the language of instructions and the language of response. For example, in one occasion the enumerator explained the missing number task using sentences such as “**Ezangi** combien?” [Missing how many?] to which the student responded “neuf” [nine], to which the enumerator responded back “Tres bien c’est neuf

que manqué” [Very good, it is nine which is missing]. The instruction was clearly translingual, and the responses later provided by that student were of the form “**Ezangi** neuf” [Missing is nine], which had some similarities with the way in which the enumerator gave the instruction. Further analyses of the tape-recordings may uncover additional relationships.

Excerpt 1. Examples of Responses to Each Task (Lingala in Bold)

Number Comparison

- **Oyo eleki** trente-huit [This (number) exceeds thirty-eight]
- **Oyo eleki** cent cinquante-quatre [This (number) exceeds one hundred and fifty four]

Missing Number

- **Awa** neuf **ezangi** [Here nine (is) missing]
- **Awa** quarante **ezangi** [Here forty (is) missing]
- **Moko** douze [The number is twelve]

Excerpt 2. Transcript. Number Comparison task (girl # 20) -Instructions (Lingala in bold)

Adult 1: Continuons. **Oyo eza nini?** [Let’s continue. Which is this number?]

Child: Neuf. [Nine]

Adult 1: **Bongo oyo?** [Then, this one?]

Child: Quatre [Four]

Adult 1: Entre les deux, **oyo eza mingi oyo wapi?** [Between both, which one is greater?]

Child: Neuf [Nine]

Adult 1: Très bien. C’est neuf. [Very good. It is nine]

4.3.3 Feedback from Teachers

To design and implement an assessment that is appropriate for the reality of Mbandaka, we need to take into account the views of teachers. First, we need to understand their views with regards on how to assess students’ knowledge, and more specifically, on how to acknowledge whether students learnt or not. The epistemic beliefs of teachers constitute a starting point from which to imagine new ways to assess students’ mathematics

proficiency. Second, we need to acknowledge the way in which teachers use evaluations. It is important to understand whether assessments are meant to inform instruction (i.e. if they are formative) or to provide a snapshot of students' achievements at particular points in time. The use that teachers give to evaluations is relevant from an epistemological standpoint and also from a practical standpoint: why are we interested in designing assessments in Mbandaka in the first place? Is it because of our own beliefs that assessments may inform instruction and may be used to evaluate the outcomes of an instructional program, or is it because teachers themselves need the information that scores convey? And why is it that they need the information? Responding to these questions is essential if we are truly interested in developing and implementing assessments that are organic to the Mbandaka reality. Last, we need to understand how teachers view and use language in their mathematics instruction. The alignment of assessments to instruction is a requirement for the valid use of test scores, and therefore, the issue of how teachers view and use language is critical.

Accordingly, I decided to focus on four themes to analyze the interviews to teachers: (a) best evaluation method – or generally, what is it that teachers believe is the correct method to assess students' performance; (b) demonstration of understanding – or how is it that teachers know whether students learnt or not; (c) use of evaluations – or what is it that teachers use the assessments' scores or information for; and (d) language of instruction – or what are the characteristics that define the use of languages in the mathematics or more generally, in the primary school classroom. The next subsections summarize my main findings.

4.3.3.1 Best Evaluation Method

Almost all the teachers mentioned traditional forms of evaluation as the best evaluation method: oral interrogation of students, homework, exercises, and/or tests. Most of the time, teachers mentioned at least two of these methods, suggesting a variety of evaluation methods. The two female teachers interviewed, mentioned that the best method was to use homework, something that was echoed by a couple of other colleagues. Most of the male teachers said that they preferred interrogations, exercises, or tests, mostly developed from the books that they had. It was unclear whether the books came from the Vas'Y Fille project or from the government, and I did not ask, but it was clear that teachers used books to extract exercises and produce assessments. One of the teachers complained of not having enough material for the students: he said that if students had their materials, it would be much easier to guide them and to assess their knowledge.

4.3.3.2 Demonstration of Understanding

Most of the teachers said that they knew their students had learnt “a knowledge” through their responses to the exercises, homework, and test. In particular, there was almost perfect consistency between their preferred method of evaluation and the way in which they realized whether students were learning or not. Teachers tended to be a little more descriptive than with regards to the previous topic, but not enough to fully capture epistemic beliefs. However, one of the teachers provided quite an interesting response (see Excerpt 4.2). According to this teacher, what made a difference between girls who learnt and girls who did not learn was their attitude, their self-confidence, apparently something that was visible in the way they spoke and wrote. This teacher used expressions such as “even in front of the crowd”, “even in front of men” which somehow indexes the Mbandaka culture.

Excerpt 4.2 Teacher explaining how she realizes when a girl learns

1 Teacher	“Je sais qu’une élève, une fille, a appris une connaissance à partir de sa
2	façon de parler, tout d’abord et d’écrire [I know that a student, that a girl
3	has learnt “a knowledge” from her way of speaking first of all and from her
4	way of writing]
5 Interviewer	D’écrire [of writing]
5 Teacher	Oui [yes]
6 Interviewer	D’accord. Est-ce que tu peux expliquer un peu plus ? Je suis très intéressée
7	de savoir qu’est qu’il change quand la fille apprend ? [Ok. Can you explain
8	a little bit more? I am very interesting in understanding what is it that
9	change when a girl learns]
10 Teacher	Bon, quand une fille apprend elle est différente avec celle qui n’est pas
11	apprennes. Les filles deviennent courageuses même devant la masse,
12	même devant les hommes il n’y a pas des complexes [Well, when a girl
13	learns she is different from those who don’t learn. The girls become braver,
14	even in front of the crowd, even in front of men there are no complexes]

4.3.3.3 Use of Evaluations

There was some incoherence between teachers in their responses with regards to how they used evaluations. All teachers used evaluations to check whether students had learnt or not, and in cases where most of the students had not learnt, some of them would repeat a lesson until they realized that students understood the content. This is interesting as some of the other teachers voiced very clearly that they could not change lessons, suggesting some type of scripted lessons plan in place. On the other hand, some teachers said that whenever students did not learn the content, they would send them to remediation; it was in remediation days that students could catch up. Many educational interventions in contexts such as DRC encompass remediation days. Teachers use remediation days to reinforce core concepts or to ensure that students learn what they are supposed to learn. Teachers typically have high degrees of flexibility in how to design and conduct remediation lessons. Therefore, while teachers did not elaborate much in terms of

responses, they suggested that assessments provided formative information, and that they used them to somehow plan and/or execute these remediation days.

Teachers also mentioned that they changed their “strategy”, mostly meaning that they changed the way in which they taught the content. This is interesting as it speaks to a self-conscious idea that teachers are responsible for children’s learning. However, there was some variation in the way in which teachers used the term strategy. In particular, one teacher used the word “strategy” to mean “focus”: to him, changing strategy meant to address the content that students should reinforce. There is not enough information to elaborate on this semantic difference, yet it is important to mention that even at this level of discourse, we noted semantic variation, and this does speak to the way in which people from Mbandaka use language.

4.3.3.4 Language of Instruction

None of the teachers said that they used one single language in the classroom. Some teachers began saying that French was the language of instruction, but that students did not fully understand French or could not fully express themselves in French, and that therefore, whenever needed, they used Lingala to explain to ensure that students understood the content. Other teachers began by saying that they used both languages for instructions, but after some iterations, changed their answer and said that while they used both French and Lingala, they typically used Lingala to ensure a better comprehension. Other teachers said that they used both languages, French and Lingala, in “pure” or mixed versions (to “mix Lingala with French”), and that students were comfortable with both languages.

We note some patterns from these responses. First, their responses suggest that all students understand Lingala, and that this is the language in which they learn better. French is the preferred language of instruction, the “recommended” language – as one school director said – yet it cannot be used in isolation, as students would not understand the content. This patterns were consistent across the responses of teachers from grades 3 to 6, pointing out that Lingala is actually used beyond its legal “recommendation”. Second, some teachers mentioned that students knew both, French and Lingala, but then contradicted themselves by adding that they could not use solely French as students would not fully understand the content. This contradiction points to a different understanding of what knowing French really means. To me, knowing French means commanding French, i.e., understanding French, i.e. not needing Lingala to support instruction. Yet to the teachers, knowing French meant “knowing some French” or being somehow familiar with French, but not commanding it.

Perhaps the most interesting aspect of the teachers’ responses was the way in which they referred to Lingala. All of them referred to French by its name, but not all teachers referred to Lingala by its name. The interviews always took place at the school director’s offices, and sometimes, school directors were present, and very few times, they intervened. In one of these interventions, the school director said that whenever there were difficulties in delivering instruction, they used “the mother tongue”, and afterwards he used the expression “local language” to restate the same. Since the teacher and myself were using the term Lingala, it was clear to all of us what these terms meant, yet it was interesting to note the difference. Another teacher once used the term “our national language” to refer to

Lingala. This reflects how the speakers of Lingala hold different beliefs towards their tongue.

4.3.4 Feedback from Enumerators

While two enumerators helped me to administer the EGMAs, and both of them were interviewed, one of the recordings was inaudible. Therefore, I only worked with one interview. Figure 5 shows the translated transcript of the interview to the enumerator. We note that the enumerator felt very comfortable implementing the translingual administration. The translingual administration was administered by the enumerators and myself, the use of language was fluid, and the whole process was a lot more interactive than usual. As per her responses, this enumerator did not seem to have any problems with the differences in administration.

To the enumerator the only difference between administrations had to do with the fact that students could access the content in two languages. There was some inconsistency in her responses but in my interpretation, to her, the most important advantage of using the translingual EGMA was that students could access the content in either French or Lingala. Neither of the girls would be damaged by a translingual administration, as both proficient French speakers and non-proficient French speakers would be able to access the content during a translingual administration. She did not have concrete suggestions to improve the assessment.

Figure 7. English Translation of the Interview to Enumerator

Interview to énumérateur

1 When you want Therese, you tell me.

2 How do you prefer to administer the EGMA, using the traditional or new method?

3 Both ways are good, the traditional way is good, and the new way is also good because there
4 are girls who understand French a little bit better so the new way is good, and then the
5 traditional way is also good, really, both are good, but I really loved the new way more than
6 the traditional way because that makes a little that that I wanted to say that that allows the girl
7 to understand something in French and in Lingala.

8 Okay, and what do you think is the main difference between the two administrations,
9 only this? That the girl understands better? Or are there other differences you want to
10 comment?

11 In this level really there are not enough differences it is only, language, because language,
12 like there is French, there are other girls who do not know how to speak French, who “listen
13 very bad French”, but when we “take” the girl in both languages, it's really better.

14 Okay, did you feel comfortable administering the EGMA in a new way?

15 Yes yes I was really comfortable, it is very good really.

16 Do you have ideas to improve the procedure of the administration of EGMA in
17 general?

18 Yes, I have ideas. Compared to, for example, eeee in relation to language, only in relation to
19 language, as in “our home” here, in almost all the schools that we passed, there are students
20 who understand a little French but many students only understand Lingala, then, this
21 administration is also good because, the girls are also able to understand French, the French
22 test.

23 Thank you very much.

CHAPTER 5

DISCUSSION

The purpose of this study was to evaluate the effectiveness and appropriateness of using a translingual administration of the EGMA assessment in measuring the mathematics proficiency of emergent bilingual girls in the region of Mbandaka, in the Democratic Republic of the Congo. By translingual administration we mean an administration of the EGMA instrument that involves more than one enumerator, that enables interactions between students and enumerators, and therefore, create spaces for translanguaging. By translingual administration we also mean an administration that uses language fluidly, both at the instructions/prompt level and at the response level. The two enumerators involved in this study used language flexibly, strategically using resources from both Lingala and French, to give instructions or explanations to the students. The students, as well, provided their answers using their entire Lingala-French linguistic repertoire. Girls did not stick to any language or any “language dosage” but rather *languaged* in a way that they could effectively use their linguistic resources to demonstrate knowledge.

The effectiveness of the administration was evaluated by means of comparing the scores of the translingual and the traditional administrations. The first research question examined whether the translingual administration had any impact on the distribution of the EGMA scores in terms of shape and means. This was evaluated both at the aggregated and at the task level. Using the average score across the baseline and midline forms, we found that the translingual administration did not significantly change the distribution of scores. However, using ANOVA and regression analysis, we learned that the effect of using a translingual administration was contingent upon the self-reported home language of the

girls. The translingual administration had a significant and positive effect on the performance of girls who reported speaking French or a mix of French and Lingala at home. Due to sampling issues, these results cannot be generalized into the larger population of “non-Lingala” speakers. Nevertheless, these outcomes suggest a relevant effect of translingual administration on bilingual girls or language minorities, that should be furthered explored.

The fact that the translingual administration improved the score of French speakers or French-Lingala speakers is thought-provoking. French is a minority language in the city of Mbandaka, since most of the daily interactions can be held in Lingala, the lingua francae (Bokamba, 2008). Different to other places in DRC, such as Kinshasa, French is a secondary language, only needed in relatively few - yet elite - spheres: for example, French is the language required in white-collar jobs or in government positions (Bokamba, 2008). The results of this work are therefore intriguing as it raises the question of who is translanguaging serving, is it the elite Mbandaka student? In other words, who is the girl who identifies herself as a French speaker in the region of Mbandaka? It could be that Mbandaka is one of those spaces where linguistic minorities are also the elite, and that translanguaging is serving kids who come from a higher socio-economic status. Such a situation neither reduces the validity nor the relevance of the findings, yet it raises unexpected unintended consequences on the equity of the system that need to be explore in more depth.

In a second look at the data, we found that among these group of girls, only one declared speaking French-only at home. In addition, the distribution of bilingual/French girls was evenly spread across sampled schools, suggesting that there is little reason to

believe that they represent a socio-economic cluster – otherwise, we would have likely found a concentration of these girls in one of the schools. Also, it is important to mention that Lingala is a powerful language in DRC: it is the language of the music, of the arts, of the armed forces, of the Catholic clergy in Kinshasa, etc. (Bokamba, 2008; 2009). Therefore, it is possible that the bilingual/French girls who were benefited by translanguaging did not belong to an elite. It could be that the French or French-Lingala speaking girls are first or second generation immigrants from neighboring countries and that they belong to marginalized communities. The question of who are the bilingual or French speaking girls should be properly explored, considering sampling issues, complementary data, the fact that there are many variations of Lingala in DRC (Bokamba, 2009), and the sociocultural complexities that exist in a context such as Mbandaka. All we can conclude from this data set and analysis, is that a translingual administration of the EGMA assessment favored linguistic minorities, mostly students who identified themselves as bilingual.

The effectiveness of the translingual administration was also evaluated in terms of its effects on the reliability estimates of the EGMA tasks and total scores. Using different methods of comparison (direct comparison, proportion of observed variance, z-transformation and z-test), the evidence clearly favored the translingual administration, as all reliability estimates were improved. Still, there are some nuances in the level of improvement when we look at the effects by task. For example, the translingual administration had almost no effect in the reliability estimates of the additions task, but it made an enormous difference for the number identification task, and a non-negligible difference on most of the other task. In particular, these differential results raise questions

about the source of the reliability improvement, as the number identification task had the lowest speech requirements, and was the easiest task as per students' scores. A possible explanation is that the higher reliability does not emerge so much from the students' ability to use both languages in their responses, but from clearer instructions/administration of the tasks. Otherwise, we would have observed a higher effect on tasks such as number comparison or missing number, which provided more opportunities to engage in speech. However, these two "wordy" tasks were harder and shorter, so it is possible that the improvements in reliability in these tasks were hindered by their higher difficulty or lower number of items. Overall, the translingual EGMA improved the reliability of the scores in a statistically significant way, and is more effective than the traditional EGMA, from a reliability standpoint. However, it is unclear how translanguaging affected reliability and we need a follow up study to fully understand the source of the reliability improvements.

Overall and based on the first two research questions, we can conclude that the translingual EGMA is effective. The translingual EGMA did not provide a lot of opportunities for students to use speech, yet it significantly improved the scores of bilingual girls without altering the scores of Lingala speakers. In other words, the translingual EGMA acted as an effective and valid accommodation, meeting the interaction hypothesis, a concept akin to the gold standard in the world of accommodations. This is somehow ironic, as conceptually speaking, translanguaging and linguistic accommodations are very different. Additionally, the translingual EGMA considerably improved the reliability estimates of the test. This outcome is a seriously interesting as it shows how language practices in assessments may account for systematic sources of measurement error. Of course, both these results require further analysis and validation, but the evidence collected

in this dissertation supports the idea of using translanguaging in mathematics assessments of bilingual students.

Additionally, a third research question examined the appropriateness of the translingual administration, via several qualitative and quantitative analyses. The first analysis showed that Mbandaka presents high barriers to standardized assessments as standardization itself seems to be an exogenous concept. Irregularity was observed in every step of the EGMA administration and in every step of this whole project, raising serious doubts about the appropriateness of standardized assessments in the region. The value of standardized assessments is rooted in the comparability of the results, but low levels of enacted standardization create room for unintended misinterpretations: results that are not truly comparable are treated as such. This is a critical issue, as standardization is central to the validity argument of tests such as the EGMA. Therefore, further research should question the extent to which educational evaluators should rely on standardization in regions like Mbandaka, and provide alternatives to deal with the problem. Last, it is important to state that the issue of standardization is not the only one that raises flags about the appropriateness of the EGMA. As noted, girls were shy and likely not used to individual oral assessments in EGMA-like settings. Further validation is needed in this regard.

A second analysis regarding the appropriateness of the translingual EGMA examined the translingual behaviors exhibited by the girls while taking the assessment. In particular, we looked at lexical patterns at the test level, at the form level, and at the task level. The analysis used three different lexical patterns: French only responses, Lingala only responses, and Mixed responses. Based on frequency analyses of these patterns, results indicated that a large majority of girls used words from more than one language to

provide their responses, at the test level. In addition, girls did not always display identical behaviors across forms: almost one third of the girls responded to the same task differently across forms. This occurred mostly for the “wordy” tasks – number comparison and missing number, which were also the tasks where girls deployed higher levels of bilingualism. Bilingualism was clearly the norm among these girls.

These results are outstanding, as the EGMA test is a low-speech test, to the extent that all the tasks could be answered with single numbers. Yet most of the girls translanguaged at the task, form, and test level. This evidence strongly indicates that in Mbandaka, any monolingual approach to testing is inappropriate, even when the language of the test is so-called a mother tongue or a native language. Double monolingual approaches would be inappropriate for the same reason, for not accounting for the natural translingual practices that these girls bring. As explained extensively in the introduction, double monolingualism is ideologically equivalent to monolingualism, and negates the discursive practices of bilinguals. Because all of the aforementioned reasons, the translingual EGMA is better suited to serve this context than the traditional EGMA.

The implications of this finding point toward exploring more rich forms of translanguaging in mathematics assessments. This translingual administration was limited because the EGMA test is low-speech and was not developed using a translingual framework: a full implementation of translanguaging in assessments requires changing the way we conceive tests from the very beginning. When doing so, it is important to keep in mind that the lexical patterns observed in this work do not seem to be random. For example, girls who used one language in the assessment, used French. In addition, among those girls who used different lexical patterns across forms, all except one, used French-only to

respond to the second form. These patterns may be a consequence of the bilingual model used in DRC, which is transitional (García, 2009), meaning that after a certain grade level, all instruction is delivered in French. Therefore, a central theme that needs to be explored, is whether there are stable linguistic patterns that should be addressed in developing translingual assessments. At the end, the purpose of translanguaging in assessment is to honor the ways in which students language, not to impose exogenous ways of flexible bilingualism.

A third analysis pertaining the issue of appropriateness was an interpretive analysis of teachers' responses to a set of semi-structured questions. This analysis shed light on the fact that teachers use various evaluation methods, and that these methods are fairly traditional in the context of formal education. Also, a majority of the interviewed teachers said that they used exercises from the instructional material that they possessed, to create homework or tests. These findings are extremely interesting because it suggests that - in this context - classroom evaluation practices may be largely influenced by the instructional materials that the Ministry or educational projects provide. Such a result has quite a direct impact for this research, as it indicates that a way in which to implement translanguaging in assessments starts by developing and providing heteroglossic instructional materials. Experiences from other countries suggest that heteroglossic materials, with diversity of discourses, of languages, and voices, can meet pedagogic expectations while counteracting separatist and hegemonic linguistic practices (Busch & Schick, 2007). Developing such material should include heteroglossic methods to assess content knowledge, a move that would certainly open up the possibilities of using translanguaging in assesement. Heteroglossic materials and examples of assessments may have a high impact in a region

like Mbandaka, where teachers use evaluations both in a formative and in a summative way. Teachers stated that they know whether students had learned the content or not, mostly through the responses to interrogations, questions, and homework. Accordingly, teachers stated that they may make course corrections to their instructional plans to ensure that students learn what they must do. If translanguaging has a positive impact in the quality of the assessment data, as this dissertation showed, providing concrete examples of translingual evaluation methods to teachers should improve assessment uses.

Overall and taking all these findings into account, we cannot state that the translingual EGMA was appropriate, and answering that question would certainly require an in-depth validity study. Indeed, our findings raise doubts about the appropriateness of standardized oral assessments that need to be addressed. However, we can state that the translingual EGMA is certainly more appropriate than the traditional EGMA for the Mbandaka context. On one hand, translanguaging seems to be a common practice in the Mbandaka classroom. Teachers declared using translingual practices in their mathematics classrooms, mainly because not all students understood French well; translanguaging is used to negotiate the understanding of students who cannot fully engage in a French-only lesson. From the way in which teachers talked about Lingala (e.g. national language, local language), it is likely that translanguaging is also used as a means to perform identities in the classroom. On the other hand, girls and enumerators translanguaged naturally during the test administration, and local enumerators did not report any difficulties administering the translingual EGMA. The translingual EGMA is better aligned with the linguistic practices of the Mbandaka girls, and therefore, is more appropriate than its traditional administration. The evidence collected in this dissertation strongly suggests that

translingual assessments are worth more exploration in the context of Mbandaka, and in contexts that face similar challenges.

An interesting inquiry that emerges from this work is whether the idea of translanguaging in assessments is related to *universal design* and how. Universal design is a term developed in the architecture field, and was originally coined as “the design of products and environments to be usable by all people, to the greatest extent possible, without the need for adaptation or specialized design” (as cited in Thompson, Johnstone, & Thurlow, 2002). The ultimate goal of universal design is to be inclusive. In particular, universal design in assessments mean to design and develop assessments that allow the widest participation of students, ensuring valid inferences and uses of scores across all students who take the test (Thompson, Johnstone, & Thurlow, 2002). Thompson et al. (2002) defined certain principles to create universally designed assessments: (a) inclusive assessment population – considering all subgroups in the development process, (b) precisely defined constructs, (c) accessible and non-biased items, (d) amenable to accommodations, (e) simple, clear, and intuitive instructions and procedures, (f) maximum readability and comprehensibility, and (g) maximum legibility. These principles should ensure that a wide variety of test takers have a fair opportunity to participate in the assessment.

Using universal design in assessments is considered best practice within the psychometric community. For example, universal design is used in the development of high-stakes large-scale assessments such as Common Core assessments developed by the Partnership for Assessment of Readiness for College and Careers (PARCC) and Smarter Balanced assessments (PARCC, 2016; Smarter Balanced Assessment Consortium &

National Center for Educational Outcomes, 2016). Both assessment systems develop their items and build their tests using a universal design approach, and complement this framework with accommodations for students who participate of certain educational programs (PARCC, 2016; Smarter Balanced Assessment Consortium & National Center for Educational Outcomes, 2016). Therefore, universal design does not replace accommodations: indeed, as noted from the fourth principle (see previous paragraph), universally designed tests need to be amenable to accommodations. The core idea behind this principle is that a large part of the construct-irrelevant variance and of the accessibility problems can and should be addressed by using universal design principles from the initial stages of test development, yet when these problems are not fully resolved, educators and test administrators may support their students through accommodations.

These ideas resonate perfectly with the common understanding of fairness in the psychometrics community. To be clear, there is no universally accepted definition of fairness in the context of testing (Ziemy, 2015), yet there is consensus around certain principles. Ziemy (2015) – who provides a summary of current perspectives on fairness and provides guidance on how to develop fair tests - states that psychometricians reject the view of fairness as equal outcomes across subgroups. Indeed, the central concern in the community is not at all about treating examinees equally but rather *equitably*, meaning that each examinee deserves an equal opportunity to participate of the assessments, even if that means changing certain characteristics. Per Ziemy (2015), the most dominant view on fairness in the psychometric community relates to validity, meaning that a fair test is one that yields valid interpretations and uses of scores for all subgroups. As such, the central role of test developers and administrators should reduce any type of construct irrelevant

variance, while ensuring that all students have a fair opportunity to demonstrate what they know and can do. This is exactly what universal design and accommodations are about: reducing construct irrelevant variance by increasing the access to the test, without changing the focal construct.

To a certain extent, the principles behind translanguaging can be considered similar or equivalent to those behind universal test design. Increasing access, developing tests with all students in mind, and ensuring fairness, are all shared values between both frameworks. According to López, et al. (2017), their implementation of translanguaging in assessments (used in this dissertation) can be regarded as using universal design principles, since translanguaging makes “items maximally comprehensible and accessible to ELs who may otherwise be excluded from showing what they know and are able to do regarding a particular content area” (p. 18). Yet there are fundamental differences between translanguaging and universal test design (UTD). Translanguaging is not so much about providing accessibility as it is to shifting our views around language. Universal test design may - and tends to be - very monolingual in their foundations and expectations. For example, the Smarter Balanced assessments –among the most sophisticated contemporary assessments – are developed using a universal design framework. Aside from the accommodations, reserved to the group of students who participate of certain educational programs, Smarter Balanced assessments include two different levels of support for what they define as English language learners (ELLs): *universal tools* and *designated supports* (Smarter Balanced Assessment Consortium & National Center for Educational Outcomes, 2016). Universal tools are access features provided to all students – ELLs or not - as part of their UTD framework. In turn, designated supports correspond to additional features

available to all students – ELLs or not - who need them, as per educators’ judgment. Universal tools include some linguistic-related features, namely: English dictionaries, English glossaries, and Thesaurus (non-embedded). Designated supports also include linguistic-related features, namely: text-to-speech features (computerized read aloud), translated glossaries (for math assessments), translated item or test directions, translated items, bilingual dictionaries, read aloud in English or Spanish (math assessments), and simplified directions (Smarter Balanced Assessment Consortium & National Center for Educational Outcomes, 2016). Besides the bilingual dictionaries, all these features are based on monolingual views around languages (one language or the other). And taken as a whole, these features are far from addressing the complex linguistic practices of bilinguals, which include using code-switching when providing explanations (Sayer, 2013) or using translingual repetition of ideas to increase understanding of the content and of languages (García, Flores, & Homonoff Woodley, 2014).

In theory, a faithful implementation of universal test design could result in the provision of translingual features in assessment, yet this is not enough to consider both frameworks equivalent. Translanguaging acknowledges that bilingual students possess one unified system of linguistic resources; the ultimate purpose of using translanguaging in assessments is to allow examinees to draw on their entire linguistic repertoires to demonstrate understanding (Lopez, et al., 2015). Translanguaging in assessments is not about creating an assessment in one language and providing “access” to students who do not master that form of language. The purpose is to develop tests that enable students to use their linguistic resources flexibly and purposefully, as in the bilingual classrooms. The

essential difference between translanguaging and universal design is that the latter may operate under monolingual assumptions of language separation.

Developing fair tests for linguistic minorities is one of the primordial challenges in test development (Schwabe, Von Davier, & Chalhoub-Deville, 2016). Translanguaging offers a new framework to think about assessments for emergent bilinguals, a framework that has received increased attention and support from researchers. Using translanguaging in developing and implementing assessments may fit the expectations around modern assessments, namely, to explicitly and intentionally address the multilingual competences that emerging bilinguals possess (García, 2009; Shohamy, 2011). Conducting additional research on the topic is a must, and we will discuss some further directions in the following paragraphs. However, in closing this section, it is important to echo the fact that translanguaging has not received enough attention and support from test developers or from policy makers. To transform the practice of assessments, and ideally of large-scale assessments, we need to observe a political will to shift from a monolingual to a multilingual approach to assessments, which is not granted. Multilingual approaches are not settled by only including bilingual students in the target and pilot populations. Implementing multilingual approaches in testing requires a bolder commitment to dissolve the complex and relatively unquestioned monolingual language practices that pervade psychometrics. Moving towards a translanguaging framework requires reconstructing our notions about language in a way that not only considers the practices, but also the views that bilinguals hold around their own language practices. And implementing successful translanguaging assessment policies requires educating users. Heteroglossic policies per se are not enough to counteract hegemonic language practices, to the extent that they may even

harm linguistic minorities if they are not coupled with guidance on how to effectively interpret and implement the policies (Mortimer, 2016). If translanguaging may improve the opportunities of bilinguals to engage in assessments and demonstrate what they know and can do, in ways that recognize their language practices and beliefs, and in ways that do not attack their identities or senses of place, it will only do so with a strong and intentional commitment on the side of researchers and practitioners.

5.1 Further Directions for Research

Multiple directions for research emerge from this work. Translanguaging in assessments is a work in process, in the sense that there are no clear, concrete, universal references on how to implement it. As such, it is important to validate the results of this study in a similar context, paying close attention to the issues of sampling and research design. A first step is to corroborate whether translingual administrations of monolingual assessments – such as the EGMA – benefit bilingual students and how. These studies must be set in a way that if we observe an improvement in scores or in reliability, we can understand its source (flexible language of instructions, higher levels of engagement, etc.). There are some parallel lines of research that could inform such studies. For example, some are investigating the use of *trialogues* in assessments of language proficiency (e.g. So, Zapata-Rivera, Cho, Luce, & Battistini, 2015). Trialogues refer to virtual conversations between one student and two virtual characters. Researchers are evaluating the extent to which triologue-based tasks are useful in assessing language proficiency, and preliminary results are positive (So, Zapata-Rivera, Cho, Luce, & Battistini, 2015). For example, So et al. (2015) show that students engage in triologue-based tasks, and generally like talking with people on the computer. Because translanguaging occurs in interaction, this line of

research may inform the implementation of interactive components in translingual assessments. This line of research can also inform how higher levels of engagement on the side of students may affect properties of scores such as reliability, among other outcomes.

A critical step to develop translingual assessments is to grasp a deep understanding of language practices and beliefs or views across different communities of students. For example, if the Mbandaka girls have stable linguistic patterns when comparing numbers, one that uses French numbers and Lingala operators/verbs, then any assessment of number comparison for that group of girls should take that practice as the norm. In addition, an assessment of number comparison should incorporate mechanisms by which girls can depart from that norm per their individual linguistic practices or preferences. When developing assessments, it is also important to also consider the girls' beliefs around languages. For example, this pattern of responses could be considered as some type of code-switching by certain linguistics or researchers, but maybe the girls do not view this practice as code-switching. Therefore, labeling this pattern as code-switching would be pointless, as the communicative and social meaning that code-switching normally conveys would be absent. In the case of mathematics, this does not seem to be extremely relevant, but for an assessment of subjects like Language or History, this occurrence could be misinterpreted as some type of rhetoric device, for example; gaining deeper insight into language practices and views should be probably best done by subject and age groups. In any case, not accounting for the linguistic views of the participants is necessarily a validity threat.

Previous research indicates that performance-based tasks or full assessments, are better for bilingual students, since they provide increased opportunities when compared to

other traditional types of assessment (García, 2009). Performance assessments are a particular type of assessment designed to measure the ability of examinees to perform certain tasks that are typically complex in nature and that require students to demonstrate the application of knowledge, skills or abilities in contexts that resemble real life situations where these are relevant (AERA, APA, & NCME, 2014; Lane & Iwatani, 2015). The promising characteristic of performance-assessments relates to the fact that they are typically administered by teachers, who may tap the bilingual resources of students while administering the assessment. Further research on translanguaging should examine the particular ways in which performance assessments activate bilingual resources and whether these characteristics are replicable with assessments of larger scale. Researchers could also examine the formative potential that performance assessments possess for emerging bilinguals. Well-designed performance assessments can communicate what are considered good models of teaching and learning (Lane & Iwatani, 2015); such outcomes would be highly beneficial for bilingual students. Future studies could therefore examine the ways in which performance assessment can be formative for students and also for teachers, who may gain deeper understanding on the language practices of students and ways by which trigger learning.

Every test of academic achievement is partly a language test (García, 2009; Schwabe, et al. 2015), because content is delivered through language. But language proficiency is a matter of important disagreement between scholars. There exist several conceptualizations of language and models of language proficiency, which go from cognitive-based models to task-centered models - or models that account for the specific contexts in which language is used (Schwabe, et al., 2015). While the topic of modelling

language proficiency is vast and definitely beyond the scope of this work, there is a direct connection between them and the way test developers think of and use language in assessments. Since translanguaging requires one to think differently about language, it is critical to modify the notions about language proficiency used in test development. Translanguaging implies focusing on meaning rather than on form. Among other things, translanguaging recognizes that what truly distinguishes multilinguals in terms of competences is not the “number of languages” that they master, but the diversity of their language experience and use (Hall, Cheng, & Carlson, 2006). *Multicompetence*, or the cognitive competence of multilinguals, develops in interaction with the social or educational environment, and multicompetent individuals are able to use their linguistic knowledge appropriately across different settings and for different purposes (Franceschini, 2008, 2011). Models of language proficiency that are appropriate for bilinguals need to incorporate social, contextual, and strategic dimensions of language use. Therefore, it is critical to expand the research around models of language proficiency used in testing, so that they recognize translanguaging.. Exploring and documenting translingual models of language proficiency is essential to ensure feasible translingual testing.

Part of the language testing field is moving towards task-based methods of assessing language proficiency (Schwabe, et al., 2015). One of the issues with these task-based or contextualized frameworks, is that there is a high degree of interaction between test takers and tasks: there is a large level of task specificity. The instability of performance across tasks in language proficiency assessments, has implications for the development of translingual assessments, which would likely use contextualized methods. The interaction between person and task entails comparability and generalizability. Therefore, a parallel

and pivotal question that emerges is whether translanguaging solutions should be local or universal. If we develop assessments based on a particular group of users, based on their language practices and beliefs, to what extent can we use these assessments with other groups of examinees? Is task-specificity something that can be solved with more robust models or a consequence of developing user-based solutions? The answer to these questions is essential to guide the future development of translanguaging in assessments and to appropriately scale it up.

A more traditional line of research that emerges from this work, is to explore innovative item formats that could enable translingual practices in assessments. Some espouse that computer-based testing (CBT) facilitates the implementation of translanguaging in assessments (e.g. López, et. al, 2017). Technology would be beneficial for test delivery, for administration, for scoring, for reporting, for creating interactions, and for enabling diverse self-regulated forms of support. The appropriateness of using CBT depends on the context where assessments are to be developed: the familiarity with computers and technology playing a crucial role in that judgment. For example, CBT is not very promising in the context of Mbandaka where girls do not have access to technology in their homes or in their schools, but it makes a lot of sense to develop CBT solutions in contexts like the US where most kids have some kind of interaction with technology in their daily lives. For technology rich contexts, 21st century technology may facilitate the implementation of translanguaging in assessments. For instance, computer-based testing may profit of top-notch technologies such as natural language processing (NLP) to analyze speech or text produced by test takers, so as to later adapt the translingual solutions to each test taker. A computer-based content assessment for bilinguals may begin with a

conversation between examinees and virtual characters in order to (a) engage examinees and to (b) understand the ways in which they use language; following these speech patterns, a software could assemble a test form, tailored in terms of linguistic practices. An important line of research is that which looks at maximizing the use of technology to develop linguistically appropriate assessments. This line of research should also explore whether it is appropriate to develop automated scoring systems for translingual assessments, one of the noticeable ways in which computer-based items improve traditional assessments. Studies on how to develop automated scoring systems under translingual frameworks, should be informed by scoring techniques that are appropriate for bilingual students, such as conceptual scoring in the case of assessing lexical command (Pearson, Fernández, & Oller, 1993). This line of research is rather complex, but it is fundamental to scale up translanguaging solutions.

Further test development using translanguaging should explore more sophisticated assessment frameworks, such as evidence-centered design, which is a principled and layered approach to test development (see for example Riconscente, Mislevy, & Corrigan, 2015). A principled- or construct-based approach to test development, that outlines the different claims we want to make of students, and how the evidence collected in translingual assessment feed these claims, may improve every aspect of the implementation. For example, in conjunction with proper analyses, principled-based frameworks may clarify which specific aspects of the translingual administration improve scores or reliability estimates. In parallel, translingual assessments need to undergo considerably validation, and the validation of the interpretation and use of scores should also follow best practices. The validation of claims, uses, and interpretations of scores

should profit from an argument-based approach to validation, which states the proposed interpretations and uses of a test explicitly, and organizes validity evidence to evaluate the assumptions and interpretations embedded in the proposed interpretation and use (Kane, 2015). An argument-based approach to validation is not only best practice but may help improve the design and development of assessments, something that translanguaging assessments - rather incipient - can profit extensively. Overall, research on how to implement translanguaging should be supported by the most advanced methods used in modern assessment.

Last, we note a huge gap between theory and practice when it comes to translanguaging. Perhaps the most urgent research agenda should be that of how to promote translanguaging assessment practices. A first step is to investigate ways by which to develop heteroglossic instructional materials, which include translanguaging exercises and assessments. Heteroglossic instructional materials may have an important impact in instruction, especially in contexts like Mbandaka where teachers have limited access to complementary resources and rely almost exclusively in their instructional books. A second step is to incorporate the language dimension in the informal and formal evaluations of assessments. For example, common alignment methodologies – or the methodologies to understand the degree to which curriculum, instruction, and assessment work together to support a common goal - do not typically include a language dimension. There are several alignment methodologies but the best known and mostly used ones (for example Webb, Achieve, or Surveys of Enacted Curriculum, to name a few) do not look at the congruence of language practices between assessment and instruction. The language component is largely absent from the alignment methodologies, yet it should be an element to consider when evaluating

the extent to which tests and instruction are mutually supportive. Validity frameworks should also consider incorporating language issues in the evidence to be gathered. Finally, we should encourage key stakeholders – teachers, principals, parents, policy makers – to continuously and increasingly incorporate heteroglossic and translingual practices in assessment matters, and educate them on how to evaluate their outcomes. No area in bilingual education is in more need than bilingual assessment (García, 2009), and the transition to appropriate assessment practices will only take place after extensive research that improves our understanding of translingual practices, our ability to develop translingual assessments, and our capacity to share and escalate effective solutions.

5.2 Limitations

Two main limitations of this study pertain to the analysis piece. On one hand, I did not sample girls using a “home language” stratification variable, nor did I randomly selected individuals based on the language spoken at home. The limitation that this entails is that I cannot generalize the results across linguistic categories. In other words, while this data shows that a translingual administration benefited bilingual girls, I cannot extend that claim beyond the sample of girls who participated in this study. On the other hand, I did not link the different audio-recordings to the EGMA data or to the background characteristics of the students. This prevented me from exploring certain results with more depth, such as the potential relationship between grade level and certain lexical patterns. It is likely that girls who responded in French only were also 5th or 6th graders, yet I could not corroborate this because of the disconnection between the different analyses.

Other limitations pertaining the analysis, although less relevant than the previous ones, refer to the test-speech analysis and to the interviews’ analysis. First, the test-speech

analysis used in this study focused exclusively on lexical characteristics, leaving out other characteristics that could have enlightened other ways of translanguaging. A concrete example of this is rooted in the accent patterns of these Congolese girls, which could also uncover different indexicalities and more subtle ways of doing translanguaging. Second, with regards to the analysis of the interview data, some pieces of the interviews were absolutely inaudible, or very difficult to transcribe. Therefore, I might have misheard or misunderstood some words or pieces of speech. Additionally, it was frequent to note grammatical deviations as per traditional standard French rules, so in many cases, it was difficult to judge whether the teachers were saying something “differently” or if I was not understanding what they were truly saying. The interpretation of the interviews was not free of ambiguity and I made my best to interpret the speakers in their own voice. However, it is important to clarify that this limitation caused minor difficulties to the interpretation, as the essential information was captured in every single interview.

Another limitation that I encountered in this work is that the EGMA test was not ideal to study translanguaging. The fact that the test could be answered only with numbers restricted the possibilities to evaluate the impact of a translingual administration of the EGMA. This limitation was further enhanced by the fact that I do not speak Lingala, and therefore, could not always participate of the dialogue between the other enumerator and the student, to either encourage using sentences to provide responses, or just to stimulate more dialogue between all the participants. The fact that I did not speak Lingala also restricted the flexible use of language, because every time I intervened I did so in French, and at certain moments the administration felt more like two monolinguals exchanging speech and providing instructions, than two enumerators using language flexibly. Last, me

as an enumerator posed a more complex problem of creating some symbolic hierarchy between the three participants for differences in race (to them, I am white) and of nationality. While I tried my best to create a comfortable environment for the girls, I could not conceal my status of foreigner. The disposition of the girls and co-enumerators might be have been different have we used a local or different “third person” in this study.

APPENDICES

APPENDIX A

EGMA BASELINE FORM

CODE 7 50 80 90 100

Enquête Midline 2015

Heure du début du test EGMA/EGMA [] heure [] minutes

Maintenant nous allons jouer quelques jeux mathématiques

Activité 1A: Identification des Nombres

Voici quelques nombres. Je veux que tu pointes du doigt chaque nombre et que tu me dises de quel nombre s'agit-il. Lire de gauche à droite, ligne par ligne. Je vais me servir de ce chronomètre et je te dirai quand commencer et quand t'arrêter. [pointez du doigt le premier chiffre] [la réponse peut être donnée en langue locale]
[instructions en langue locale]

Commence par toi. C'est quoi, ce nombre?
[instructions en langue locale]

(/) Incorrecte ou pas de réponse () Après dernier nombre lu

7	9	0	13	40	(5)
68	45	29	43	98	(10)
31	55	64	37	85	(15)
106	345	587	831	959	(20)

Feuille A

60 secondes. Si le temps sur le chronomètre est épuisé arrêter l'activité

Si l'enfant s'arrête sur un nombre pendant 5 SECONDES

Temps restant:

Crochet:

Nombre (s) correct(s):

Activité 2A: Comparaison des Quantités - EXEMPLE

Regarde ces nombres. Dis-moi lequel est le plus grand.
[instructions en langue locale]

- Si l'élève répond correctement, dites-lui : **C'est correcte, 9 est plus grand..**
[instructions en langue locale]
- Si l'élève ne répond pas correctement, dites-lui : **Le nombre 9 est plus grand. Ce nombre est 4. Ce nombre est 9. 9 est plus grand que 4. Continuons.**
[instructions en langue locale]

Feuille B1

Activité 2A : Comparaison des Quantités - EXERCICE

Regarde ces nombres. Dis-moi lequel est le plus grand.
[instructions en langue locale]
 [Pointez du doigt les nombres de chaque ligne, une par une, en répétant la consigne] [la fille doit dire le nombre en langue locale ou en français]

x sur 1 si « Correct » sur 0 = « Incorrect » ou « pas de réponse »

5	3	<u>5</u>	<input type="checkbox"/>	<input type="checkbox"/>	93	76	<u>93</u>	<input type="checkbox"/>	<input type="checkbox"/>
11	24	<u>24</u>	<input type="checkbox"/>	<input type="checkbox"/>	147	154	<u>154</u>	<input type="checkbox"/>	<input type="checkbox"/>
38	24	<u>38</u>	<input type="checkbox"/>	<input type="checkbox"/>	287	534	<u>534</u>	<input type="checkbox"/>	<input type="checkbox"/>
57	48	<u>57</u>	<input type="checkbox"/>	<input type="checkbox"/>	634	643	<u>643</u>	<input type="checkbox"/>	<input type="checkbox"/>
63	65	<u>65</u>	<input type="checkbox"/>	<input type="checkbox"/>	865	963	<u>963</u>	<input type="checkbox"/>	<input type="checkbox"/>

B2

Si l'enfant fait 4 erreurs successives

Si l'enfant ne répond pas après 5 SECONDES.

Total Correct:

Activité 3A: Chiffre Manquant - EXEMPLE

Voici quelques nombres. Un, deux, trois, quatre. Quel est le nombre qui va ici ?

C1

[Instructions en langue locale]

[pointez du doigt à l'espace]

Si l'élève répond correctement, dites-lui : **C'est juste, cinq. Continuons.**

[Instructions en langue locale]

Si l'élève ne répond pas correctement, dites-lui : **Ici, le nombre est cinq. Compte avec moi.** [Pointez chaque nombre du doigt] **Un, deux, trois, quatre. Ici, le nombre est cinq. Continuons.**

[Instructions en langue locale]

Voici quelques nombres. Dix-sept, [pointez du doigt à l'espace sans rien dire], dix-neuf, vingt, vingt-et-un. Quel nombre va ici? [pointez du doigt à l'espace]

[Instructions en langue locale]

Si l'élève répond correctement, dites-lui : **C'est juste, dix-huit.**

Si l'élève ne répond pas correctement, dites-lui : **Ici, le nombre est dix-huit.**

Compte avec moi. [Pointez chaque nombre du doigt] **Dix-sept, dix-huit, dix-neuf, vingt, vingt-et-un. Continuons.**

[Instructions en langue locale]

Activité 3A: Chiffre Manquant - EXERCICE

Voici quelques nombres. Quel est le nombre qui va ici ?

C2

[Instructions en langue locale]

[Pointez l'espace du doigt] [Répétez pour chaque item si nécessaire]

[la réponse peut être donnée en langue locale]

x sur 1 si « Correct »

x sur 0 = « Incorrect » ou « pas de réponse »

6	7	8	<u>9</u>	<input type="checkbox"/>	<input type="checkbox"/>	248	249	<u>250</u>	251	<input type="checkbox"/>	<input type="checkbox"/>
13	14	<u>15</u>	16	<input type="checkbox"/>	<input type="checkbox"/>	48	<u>49</u>	44	42	<input type="checkbox"/>	<input type="checkbox"/>
30	<u>40</u>	50	60	<input type="checkbox"/>	<input type="checkbox"/>	20	25	<u>30</u>	35	<input type="checkbox"/>	<input type="checkbox"/>
<u>400</u>	500	600	700	<input type="checkbox"/>	<input type="checkbox"/>	650	640	630	<u>620</u>	<input type="checkbox"/>	<input type="checkbox"/>
2	4	6	<u>8</u>	<input type="checkbox"/>	<input type="checkbox"/>	3	8	<u>13</u>	18	<input type="checkbox"/>	<input type="checkbox"/>



Si l'enfant fait **4 erreurs successives**

Si l'enfant ne répond pas après **5 SECONDES.**

Total correct:

x

Activité 4A: Addition

Voici quelques additions. Je vais maintenant utiliser ce chronomètre. Escale de ton mieux. Si tu ne connais pas une réponse, passe au prochain problème. Es-tu prêt(e) ? ... Commence par toi et va ligne par ligne.

[instructions en langue locale]

[instructions en langue locale]

[Pointez le premier item] [la réponse peut être donnée en langue locale]

(/) Incorrecte ou pas de réponse () Après dernier item répondu


$1 + 4 = (5)$	$2 + 3 = (5)$	$5 + 3 = (8)$	(3)
$3 + 4 = (7)$	$3 + 3 = (6)$	$7 + 1 = (8)$	(6)
$8 + 2 = (10)$	$2 + 9 = (11)$	$4 + 6 = (10)$	(9)
$9 + 2 = (11)$	$6 + 7 = (13)$	$4 + 8 = (12)$	(12)
$8 + 6 = (14)$	$7 + 4 = (11)$	$9 + 8 = (17)$	(15)
$5 + 6 = (11)$	$8 + 8 = (16)$	$9 + 5 = (14)$	(18)
$10 + 3 = (13)$	$9 + 10 = (19)$		(20)


Temps restants

Crochet

Total Correct:

 D1

 60 secondes, Si le temps sur le chronomètre est épuisé arrêter l'activité

 Si l'enfant s'arrête sur un nombre pendant **5 SECONDES**

Activité 5A: Soustraction

Voici quelques soustractions. Je vais encore utiliser ce chronomètre. Escale de ton mieux. Si tu ne connais pas une réponse, passe au prochain problème. Es-tu prêt(e) ? ... Commence par toi et va ligne par ligne.

[instructions en langue locale]

[instructions en langue locale]

[Pointez le premier item] [la réponse peut être donnée en langue locale]

(/) Incorrecte ou pas de réponse () Après dernier item répondu


$5 - 4 = (1)$	$5 - 3 = (2)$	$8 - 3 = (5)$	(3)
$7 - 4 = (3)$	$6 - 3 = (3)$	$8 - 1 = (7)$	(6)
$10 - 2 = (8)$	$11 - 9 = (2)$	$10 - 6 = (4)$	(9)
$11 - 2 = (9)$	$13 - 7 = (6)$	$12 - 8 = (4)$	(12)
$14 - 6 = (8)$	$11 - 4 = (7)$	$17 - 8 = (9)$	(15)
$11 - 6 = (5)$	$16 - 8 = (8)$	$14 - 5 = (9)$	(18)
$13 - 3 = (10)$	$19 - 10 = (9)$		(20)


Temps restants

Crochet

Total Correct:

 D2

 60 secondes, Si le temps sur le chronomètre est épuisé arrêter l'activité

 Si l'enfant s'arrête sur un nombre pendant **5 SECONDES**

APPENDIX B

EGMA MIDLINE FORM

CODE P SD RX E Classe <u>Elève</u>																																																			
Heure du début du test [] heure [] minutes ☛ <i>Maintenant nous allons jouer quelques jeux mathématiques</i>																																																			
Activité 1A: Identification des Nombres																																																			
☛ Voici quelques nombres. Je veux que tu pointes du doigt chaque nombre et que tu me dises de quel nombre s'agit-il. Lire de gauche à droite, ligne par ligne. Je vais me servir de ce chronomètre et je te dirai quand commencer et quand t'arrêter. [pointez du doigt le premier chiffre] [Instructions en langue locale]	📄 Feuille A ⌚ 60 secondes, Si le temps sur le chronomètre est épuisé arrêter l'activité ➡ Si l'enfant s'arrête sur un nombre pendant 5 SECONDES																																																		
☛ Commence par ici. C'est quoi, ce nombre? [Instructions en langue locale]																																																			
☒ (/) Incorrecte ou pas de réponse () Après dernier nombre lu <table style="width: 100%; border-collapse: collapse; text-align: center;"> <tr> <td colspan="5">Cum.</td> <td colspan="1">Tot.</td> </tr> <tr> <td style="border: 1px solid black;">2</td> <td style="border: 1px solid black;">9</td> <td style="border: 1px solid black;">0</td> <td style="border: 1px solid black;">22</td> <td style="border: 1px solid black;">16</td> <td style="border: 1px solid black;">(5)</td> </tr> <tr> <td style="border: 1px solid black;">35</td> <td style="border: 1px solid black;">12</td> <td style="border: 1px solid black;">28</td> <td style="border: 1px solid black;">60</td> <td style="border: 1px solid black;">72</td> <td style="border: 1px solid black;">(10)</td> </tr> <tr> <td style="border: 1px solid black;">67</td> <td style="border: 1px solid black;">89</td> <td style="border: 1px solid black;">39</td> <td style="border: 1px solid black;">97</td> <td style="border: 1px solid black;">55</td> <td style="border: 1px solid black;">(15)</td> </tr> <tr> <td style="border: 1px solid black;">234</td> <td style="border: 1px solid black;">107</td> <td style="border: 1px solid black;">692</td> <td style="border: 1px solid black;">528</td> <td style="border: 1px solid black;">912</td> <td style="border: 1px solid black;">(20)</td> </tr> </table>	Cum.					Tot.	2	9	0	22	16	(5)	35	12	28	60	72	(10)	67	89	39	97	55	(15)	234	107	692	528	912	(20)																					
Cum.					Tot.																																														
2	9	0	22	16	(5)																																														
35	12	28	60	72	(10)																																														
67	89	39	97	55	(15)																																														
234	107	692	528	912	(20)																																														
Temps restant:	☒																																																		
Crochet:	☒																																																		
Nombre (s) correct(s):	☒																																																		
Activité 2A: Comparaison des Quantités - EXEMPLE																																																			
☛ Regarde ces nombres. Dis-moi lequel est le plus grand. [Instructions en langue locale] <ul style="list-style-type: none"> • Si l'élève répond <u>correctement</u>, dites-lui : ☛ C'est correcte, 9 est plus grand.. • Si l'élève ne répond <u>pas correctement</u>, dites-lui ☛ Le nombre 9 est plus grand. Ce nombre est 4. Ce nombre est 9. 9 est plus grand que 4. Continuons. [Instructions en langue locale]	📄 Feuille B1																																																		
Activité 2A : Comparaison des Quantités - EXERCICE																																																			
☛ Regarde ces nombres. Dis-moi lequel est le plus grand. [Instructions en langue locale] [Pointez du doigt les nombres de chaque ligne, une par une, en répétant la consigne]	📄 B2 ✖ Si l'enfant fait 4 erreurs successives ➡ Si l'enfant ne répond pas après 5 SECONDES .																																																		
☒ * sur 1 si « Correct » * sur 0 = « Incorrect » ou « pas de réponse » <table style="width: 100%; border-collapse: collapse; text-align: center;"> <tr> <td style="border: 1px solid black;">8</td> <td style="border: 1px solid black;">5</td> <td style="border: 1px solid black;"><u>8</u></td> <td style="border: 1px solid black;">1</td> <td style="border: 1px solid black;">0</td> <td style="border: 1px solid black;">94</td> <td style="border: 1px solid black;">77</td> <td style="border: 1px solid black;"><u>94</u></td> <td style="border: 1px solid black;">1</td> <td style="border: 1px solid black;">0</td> </tr> <tr> <td style="border: 1px solid black;">15</td> <td style="border: 1px solid black;">26</td> <td style="border: 1px solid black;"><u>26</u></td> <td style="border: 1px solid black;">1</td> <td style="border: 1px solid black;">0</td> <td style="border: 1px solid black;">357</td> <td style="border: 1px solid black;">319</td> <td style="border: 1px solid black;"><u>357</u></td> <td style="border: 1px solid black;">1</td> <td style="border: 1px solid black;">0</td> </tr> <tr> <td style="border: 1px solid black;">27</td> <td style="border: 1px solid black;">34</td> <td style="border: 1px solid black;"><u>34</u></td> <td style="border: 1px solid black;">1</td> <td style="border: 1px solid black;">0</td> <td style="border: 1px solid black;">323</td> <td style="border: 1px solid black;">443</td> <td style="border: 1px solid black;"><u>443</u></td> <td style="border: 1px solid black;">1</td> <td style="border: 1px solid black;">0</td> </tr> <tr> <td style="border: 1px solid black;">40</td> <td style="border: 1px solid black;">58</td> <td style="border: 1px solid black;"><u>58</u></td> <td style="border: 1px solid black;">1</td> <td style="border: 1px solid black;">0</td> <td style="border: 1px solid black;">708</td> <td style="border: 1px solid black;">780</td> <td style="border: 1px solid black;"><u>780</u></td> <td style="border: 1px solid black;">1</td> <td style="border: 1px solid black;">0</td> </tr> <tr> <td style="border: 1px solid black;">78</td> <td style="border: 1px solid black;">75</td> <td style="border: 1px solid black;"><u>78</u></td> <td style="border: 1px solid black;">1</td> <td style="border: 1px solid black;">0</td> <td style="border: 1px solid black;">955</td> <td style="border: 1px solid black;">958</td> <td style="border: 1px solid black;"><u>958</u></td> <td style="border: 1px solid black;">1</td> <td style="border: 1px solid black;">0</td> </tr> </table>	8	5	<u>8</u>	1	0	94	77	<u>94</u>	1	0	15	26	<u>26</u>	1	0	357	319	<u>357</u>	1	0	27	34	<u>34</u>	1	0	323	443	<u>443</u>	1	0	40	58	<u>58</u>	1	0	708	780	<u>780</u>	1	0	78	75	<u>78</u>	1	0	955	958	<u>958</u>	1	0	
8	5	<u>8</u>	1	0	94	77	<u>94</u>	1	0																																										
15	26	<u>26</u>	1	0	357	319	<u>357</u>	1	0																																										
27	34	<u>34</u>	1	0	323	443	<u>443</u>	1	0																																										
40	58	<u>58</u>	1	0	708	780	<u>780</u>	1	0																																										
78	75	<u>78</u>	1	0	955	958	<u>958</u>	1	0																																										
Total Correct:	☒																																																		

Activité 3A: Chiffre Manquant - EXEMPLE

👁️ Voici quelques nombres. Un, deux, trois, quatre. Quel est le nombre qui va ici ?

📖 C1

[Instructions en langue locale]

[pointez du doigt à l'espace]

Si l'élève répond correctement, dites-lui : 👁️ **C'est juste, cinq. Continuons.**

[Instructions en langue locale]

Si l'élève ne répond pas correctement, dites-lui 👁️ **Ici, le nombre est cinq. Compte avec moi.** [Pointez chaque nombre du doigt] **Un, deux, trois, quatre. Ici, le nombre est cinq. Continuons.**

[Instructions en langue locale]

👁️ Voici quelques nombres. Dix-sept, [pointez du doigt à l'espace sans rien dire], dix-neuf, vingt, vingt-et-un. Quel nombre va ici? [pointez du doigt à l'espace]

[Instructions en langue locale]

Si l'élève répond correctement, dites-lui : 👁️ **C'est juste, dix-huit.**

Si l'élève ne répond pas correctement, dites-lui 👁️ **Ici, le nombre est dix-huit. Compte avec moi.** [Pointez chaque nombre du doigt] **Dix-sept, dix-huit, dix-neuf, vingt, vingt-et-un. Continuons.**

[Instructions en langue locale]

Activité 3A: Chiffre Manquant - EXERCICE

👁️ Voici quelques nombres. Quel nombre va ici ?

📖 C2

[Instructions en langue locale]

[Pointez l'espace du doigt] [Répétez pour chaque item si nécessaire]

👁️ * sur 1 si « Correct »

* sur 0 = « Incorrect » ou « pas de réponse »

6	7	8	9	1	0	347	348	349	350	1	0
11	12	13	14	1	0	39	37	35	33	1	0
49	59	69	79	1	0	50	55	60	65	1	0
100	200	300	400	1	0	440	430	420	410	1	0
12	14	16	18	1	0	3	8	13	18	1	0

⊗

👁️ Si l'enfant fait 4 erreurs successives

👁️ Si l'enfant ne répond pas après 5 SECONDES.

Total correct: 📖

Activité 4A: Addition

🕒 **Voici quelques additions. Je vais maintenant utiliser ce chronomètre. Essaie de ton mieux. Si tu ne connais pas une réponse, passe au prochain problème. Es-tu prêt(e) ? ... Commence par ici et va ligne par ligne.**
 [Instructions en langue locale]
 [Instructions en langue locale]

🕒 D1

🕒 60 secondes, Si le temps sur le chronomètre est épuisé arrêter l'activité

[Pointez le premier item]

⚡ (/) Incorrecte ou pas de réponse () Après dernier item répondu

1 + 1 = (2)	1 + 2 = (3)	3 + 2 = (5)	(3)
2 + 4 = (6)	3 + 3 = (6)	8 + 1 = (9)	(6)
3 + 6 = (9)	7 + 3 = (10)	5 + 5 = (10)	(9)
2 + 8 = (10)	7 + 4 = (11)	8 + 6 = (14)	(12)
5 + 7 = (12)	6 + 6 = (12)	7 + 8 = (15)	(15)
9 + 7 = (16)	8 + 8 = (16)	4 + 7 = (11)	(18)
10 + 3 = (13)	10 + 7 = (17)		(20)

⚡ Si l'enfant s'arrête sur un nombre pendant **5 SECONDES**

Temps restants

⚡

Crochet

⚡

Total Correct:

⚡

Activité 5A: Soustraction

🕒 **Voici quelques soustractions. Je vais encore utiliser ce chronomètre. Essaie de ton mieux. Si tu ne connais pas une réponse, passe au prochain problème. Es-tu prêt(e) ? ... Commence par ici et va ligne par ligne.**
 [Instructions en langue locale]
 [Instructions en langue locale]

🕒 D2

🕒 60 secondes, Si le temps sur le chronomètre est épuisé arrêter l'activité

[Pointez le premier item]

⚡ (/) Incorrecte ou pas de réponse () Après dernier item répondu

2 - 1 = (1)	3 - 2 = (1)	5 - 3 = (2)	(3)
6 - 2 = (4)	6 - 3 = (3)	9 - 1 = (8)	(6)
9 - 3 = (6)	10 - 3 = (7)	10 - 5 = (5)	(9)
10 - 2 = (8)	11 - 7 = (4)	14 - 8 = (6)	(12)
12 - 5 = (7)	12 - 6 = (6)	15 - 7 = (8)	(15)
16 - 7 = (9)	16 - 8 = (8)	11 - 7 = (4)	(18)
13 - 10 = (3)	17 - 7 = (10)		(20)

⚡ Si l'enfant s'arrête sur un nombre pendant **5 SECONDES**

Temps restants

⚡

Crochet

⚡

Total Correct:

⚡

REFERENCES

- Abedi, J. (2004). The No Child Left Behind act and English language learners: Assessment and accountability issues. *Educational Researcher*, 33(1), 4-14.
- Abedi, J. (2006). Language issues in item development. In S. M. Downing, & T. M. Haladyna, *Handbook of test development* (pp. 377-398). New Jersey, NJ: Lawrence Erlbaum Associates, Publishers.
- Abedi, J. (2009). Computer testing as a form of accommodation for English language learners. *Educational Assessment*, 14, 195-211.
- Abedi, J., Courtney, M., Mirocha, J., Leon, S., & Goldberg, J. (2001). Language accommodation for large-scale assessments in science. *National Center for Research on Evaluation, Standards, and Student Testing*.
- Abedi, J., & Ewers, N. (2013). *Accommodations for English learners and students with disabilities: A research based decision algorithm*. Smarter Balanced.
- Abedi, J., & Gandara, P. (2006). Performance of English language learners as a subgroup in large-scale assessment: Interaction of research and policy. *Educational Measurement: Issues and Practice*, Winter, 2006, 36-46.
- Abedi, J., & Liguanti, R. (2012). *Issues and Opportunities in Improving the Quality of Large Scale Assessment Systems for ELLs*. Retrieved from <http://ell.stanford.edu/publication/issues-and-opportunities-improving-quality-large-scale-assessment-systems-ells>
- Abedi, J., & Lord, C. (2001). The language factor in mathematics tests. *Applied Measurement in Education*, 14(3), 219-234.
- Abedi, J., Courtney, M., & Leon, S. (2003). *Research-supported accommodation for English language learners in NAEP*. CSE Technical Report.
- Abedi, J., Hofstetter, C. H., & Lord, C. (2004). Assessment accommodations for English language learners: Implications for policy-based empirical research. *Review of Educational Research*, 74 (1), 1-28.
- Abedi, J., Lord, C., & Plummer, J. R. (1997). *Final report of language background as a variable in NAEP mathematics performance*. Los Angeles, California: Center for the Study of Evaluation, National Center for Research on Evaluation, Standards, and Student Testing, Graduate School of Education & Information Studies, University of California.
- Abedi, J., Lord, C., Hofstetter, C., & Baker, E. (2000). Impact of accommodation strategies on English language learners' test performance. *Educational Measurement: Issues and Practice*, 19(3), 16-26.

- Adams, T. (2003). Reading mathematics: More than words can say. *The Reading Teacher*, 56(8), 786-795.
- Adoni, M., & Quing, Y. (2014). Language, Mathematics, and English language learners. *Australian Mathematics Teacher*, 70(3), 3-13.
- Albus, D., & Thurlow, M. L. (2008). Accommodating students with disabilities on state English language proficiency assessments. *Assessment for Effective Intervention*, 156-166.
- American Educational Research Association, American Psychological Association, National Council of Measurement in Education [AERA/APA/NCME]. (2014). *Standards for educational and psychological testing*. Washington, DC: American Psychological Association.
- Attali, Y. (2004) Reliability of Speeded Number-Right Multiple-Choice Tests. ETS Research Report Series, 2004 (1).
- Bailey, A. L. (2007). *The language demands of schools: Putting academic English to the test*. New Haven, USA: Yale University Press.
- Bailey, A. L., & Huang, B. H. (2011). Do current English language development/proficiency standards reflect the English needed for success in school? *Language Testing*, 28(3), 343-365.
- Baker, C. (2011). *Foundations of bilingual education and bilingualism*. Bristol, UK: Multilingual Matters.
- Blackledge, A., & Creese, A. (2014). Heteroglossia as practice and pedagogy. In *Heteroglossia as practice and pedagogy* (pp. 1-20). Springer Netherlands.
- Block, D. (2008). Language education and globalization. In N. H. Hornberger, *Language policy and political issues in education* (pp. 15-30). New York, NY: Springer Science + Business Media, LLC.
- Bokamba, E.G. (2008). The lives of local and regional Congolese languages in globalized linguistic markets. In *Globalization and language vitality: Perspectives from Africa*
- Bokamba, E. G. (2009). The spread of Lingala as a lingua franca in the Congo basin. In McLaughlin, F. *The languages of urban Africa* (pp 50-70). New York, USA: Continuum.
- Brown, T. (2002). *Mathematics education and language: Interpreting hermeneutics and post-structuralism*. Boston, USA: Kluwer Academic Publishers.
- Busch, B., & Schick, J. (2015). Educational materials reflecting heteroglossia: Disinventing ethnolinguistic differences in Bosnia-Herzegovina. In Makoni, S. &

- Pennycook, A., *Disinventing and Reconstituting Languages* (216-232). Multilingual Matters.
- Canagarajah, S. (2011). Codemeshing in academic writing: Identifying teachable strategies of translanguaging. *The Modern Language Journal*, 95 (iii), 401-417.
- Central Intelligence Agency. (2015, May 1). *The world factbook: Democratic Republic of the Congo*. Retrieved from The world factbook: <https://www.cia.gov/library/publications/the-world-factbook/geos/cg.html>
- Creese, A., & Blackledge, A. (2010). Translanguaging in the bilingual classroom: A pedagogy for learning and teaching? *The Modern Language Journal*, 94(i), 103-115.
- Cortina, J.M. (1993). What is coefficient alpha? An examination of theory and applications. *Journal of applied psychology*, 78(1), 98.
- Cronbach, L., & Meehl, P. E. (1955). Construct Validity in Psychological Test. *Psychological Bulletin*, 52, 281-302.
- Cummins, J. (2000). *Language, power and pedagogy: Bilingual children in the crossfire*. Clevedon, England: Multilingual Matters.
- de Saint Moulin, L., & Tshibanda, J. L. (2005). *Atlas de l'organisation administrative de la République démocratique du Congo:(fonds de plan de l'Institut géographique du Congo)*. Centre d'études pour l'action sociale.
- DiCerbo, P. A., Anstrom, K. A., Baker, L. L., & Rivera, C. (2014). A review of the literature on teaching academic English to English language learners. *Review of Educational Research*, 84 (3), 446-482.
- Edinga, K. (1999). La politique de l'éducation en république du Congo (Zaire) de 1960 a 1989.
- Ercikan, K., Chen, M. Y., Lyons-Thomas, J., Goodrich, S., Sandilands, D., Roth, W.-M., & Simon, M. (2015). Reading proficiency and comparability of mathematics and science scores for students from English and non-English backgrounds: An international perspective. *International Journal of Testing*, 15, 153-175.
- Escamilla, K. (2006). Monolingual assessment and emerging bilinguals: A case study in the US. In O. Garcia, T. Skutnabb-Kangas, & M. E. Torres-Guzman, *Imagining multilingual schools* (pp. 184-199). New York and Tronninge Mose: Multilingual Matters LTD.
- Franceschini, R. (2008). *Research area report, Thematic area C: Multilingualism and education (D9)*. Languages in a Network of European Excellence. Retrieved from <http://linee.info/downloads/project-contents.html?PHPSESSID=c68c233dff2499b3cae385d8f48c71a>

- Franceschini, R. (2011). Multilingualism and multicompetence: A conceptual view. *The Modern Language Journal*, 344-355.
- Gall, M. D., Gall, J. P., & Borg, W. R. (2007). *Educational research: An introduction*. (8. edition, Ed.) Boston, MA: Allyn and Bacon.
- García, O. (2007). Foreword. In S. Makoni, & A. Pennycook, *Disinventing and reconstituting languages* (pp. xi-xv). Clevedon, UK: Multilingual Matters.
- García, O. (2009). *Bilingual education in the 21st Century: A global perspective*. Malden, MA: Wiley-Blackwell.
- García, O., Kleifgen, J. A., & Falchi, L. (2008). *From English language learners to emergent bilinguals*. New York, NY: Campaign for Educational Equity.
- García, O., Skutnabb-Kangas, T., & Torres-Guzman, M. E. (2006). Weaving spaces and (de)constructing ways for multilingual schools: The actual and the imagined. In O. García, T. Skutnabb-Kangas, & M. E. Torres-Guzman, *Imagining multilingual schools: Languages in education and glo(c)alization* (pp. 3-50). García, Ofelia; Skutnabb-Kangas, Tove; Torres-Guzman, Maria E.; Multilingual matters ltd.
- Haag, N., Heppt, B., & Roppelt, A. (2015). Linguistic simplification of mathematics items: Effects for language minority students in Germany. *European Journal of Psychology in Education*, 30, 145-167.
- Hall, J. K., Cheng, A., & Carlson, M. T. (2006). Reconceptualizing multicompetence as a theory of language knowledge. *Applied Linguistics*, 27, 220-240.
- Horner, B., Lu, M.-Z., Jones Royster, J., & Trimbur, J. (2011). Language difference in writing: Towards a translingual approach. *College English*, 73(3), 303-320.
- Hudson, R. A. (1996). *Sociolinguistics*. Cambridge University Press.
- Kane, M. (2015). Validation strategies: delineating and validating proposed interpretations and uses of test scores. In S. Lane, M. R. Raymond, & T. M. Haladyna, *Handbook of test development* (2nd ed., pp. 64-80). New York, USA: Routledge.
- Kettler, R. (2015). Adaptations and access to assessments of Common Core content. *Review of Research in Education*, 39, 295-330.
- Kieffer, M. J., Lesaux, N. K., Rivera, M., & Francis, D. J. (2009). Accommodations for English language learners taking large-scale assessments: A meta-analysis on effectiveness and validity. *Review of Educational Research*, 79 (3), 1168-1201.
- Kohn, M. (2014). Colonialism. In Zalta, E.N., *The Stanford Encyclopedia of Philosophy*, Retrieved from <https://plato.stanford.edu/archives/spr2014/entries/colonialism/>.

- Kopriva, R. (2000). *Ensuring accuracy in testing for English language learners*. Washington, D.C., USA: Council of Chief State School Officers.
- L'ASBL Investing In People (IIP). (2014). *Enseignement des mathématiques au primaire*. Kinshasa, DRC: UNICEF.
- Lane, S., & Iwatani, E. (2015). Design of Performance Assessments in Education. In S. Lane, M. R. Raymond, & T. M. Haladyna, *Handbook of test development* (2nd ed., pp. 274-293). New York, USA: Routledge.
- Li, H., & Suen, H. K. (2012). The effects of test accommodations for English language learners: A meta-analysis. *Applied Measurement in Education*, 25(4), 327-346.
- Lindstrom, J. H. (2010). Mathematics assessment accommodations: Implications of differential boost for students with learning disabilities. *Intervention in School and Clinic*, 46(1), 5-12.
- Logan-Terry, A., & Wright, L. J. (2010). Making thinking visible: An analysis of ELLs' interactions with access-based science assessment items. *AccELLerate!*, 2(4), 11-14.
- López, A. A., Turkan, S., & Guzman-Orth, D. A. (2017). *Conceptualizing the use of translanguaging in initial content assessments for newly arrived emergent bilingual students*. (ETS Research Report No. RR-17-07). Princeton, NJ: Educational Testing Service.
- Mackey, W. (1972). *Bilingual education in a bilingual school*. Newbury House.
- Makoni, S., & Mashiri, P. (2006). Critical historiography: Does language planning in Africa need a construct of language as part of its theoretical apparatus? In S. Makoni, & A. Pennycook, *Disinventing and reconstituting languages* (pp 62-89). Clevedon, GB: Multilingual Matters.
- Makoni, S., & Pennycook, A. (2006). Disinventing and reconstituting languages. In S. Makoni, & A. Pennycook, *Disinventing and reconstituting languages* (pp. 1-41). Clevedon, GB: Multilingual Matters.
- Martiniello, M. (2008). Language and the performance of English-language learners in math word problems. *Harvard Education Review*, 78 (2), 333-368.
- Martone, A., & Sireci, S. G. (2009). Evaluating alignment between curriculum, assessments, and instruction. *Review of Educational Research*, 4, 1332-1361.
- 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. 381-423). Berlin ; New York: Mouton de Gruyter.

- Messick, S. (1989). Validity. In R. L. Linn, *Educational Measurement* (pp. 13-103). New York: American Council of Education and Macmillan.
- Miller, E. R., Okum, I., Sinai, R., & Miller, K. S. (1999). *A study of the English language readiness of limited English proficient students to participate in New Jersey's statewide assessment system*. Montreal, Canada: In annual meeting of the National Council of Measurement in Education.
- Mortimer, K. (2016). A potentially heteroglossic policy becomes monoglossic in context: An ethnographic analysis of Paraguayan bilingual education policy. *Anthropology & Education Quarterly*, 47(4), 349-365.
- Mukala-Missumbi, S. (2012). *La situation des jeunes filles dans les sciences et les mathématiques en République Démocratique du Congo*. Green Light for Girls.
- National Governors Association; Council of Chief State School Officers. (2012). *Common Core Standards State Initiative: Common Core Standards: Mathematics*. Retrieved from <http://www.corestandards.org/Math/>
- Nthawakuderwa, B. (1985). Authenticity and problems of language planning in Zaire. *Language in education in Africa* (pp. 101-140). Edinburgh: University of Edinburgh.
- Nunan, D. (1992). An introduction to research methods and traditions. In D. Nunan, *Research methods in language learning* (pp. 1-23). Cambridge, UK: Cambridge University Press.
- Olivieri, M. E., Ercikan, K., & Simon, M. (2015). A framework for developing comparable multilingual assessments for minority populations: Why context matters. *International Journal of Testing*, 15(2), 94-113.
- Partnership for Assessment of Readiness for College and Career. (2016). *PARCC Accessibility Features and Accommodations Manual*. PARCC. Retrieved from <http://www.parcconline.org/assessments/accessibility>
- Pennock-Román, M. (2002). Relative effects of English proficiency on general admissions tests versus subject tests. *Research in Higher Education*, 43 (5), 601-623.
- Platas, L. M., Ketterlin-Gellar, L., Brombacher, A., & Stiabkhan, Y. (2014). *Early Grade Mathematics Assessment (EGMA) Toolkit*. RTI International.
- Press, W. H., Teukolsky, S. A., Vetterling, W. T., & Flannery, B. P. (1992). Numerical recipes in C: The art of scientific computing. Cambridge, UK.
- Randall, J. (2015). Girls enrolling, persisting, and achieving: Impact findings of Vas-Y-Filles in the Democratic Republic of the Congo. Oxford, UK: Paper presented at the UK Forum for International Education and Training Biannual Meeting.

- Reubens, A. (2009). *Early grade mathematics assessment (EGMA): A conceptual framework based on mathematics skills development in children*. USAID. Retrieved from <https://www.eddataglobal.org/documents/index.cfm?fuseaction=pubDetail&id=193>
- Riconscente, M., M., Mislevy, R., J., & Corrigan, S. (2015). Evidence-centered design. In S. Lane, M. Raymond, & T. Haladyna, *Handbook of test development* (2nd ed., pp. 81-99). New York, USA: Routledge.
- Rivera, C., & Stansfield, C. W. (2001). Leveling the playing field for English language learners: Increasing participation in state and local assessments through accommodations. In R. Brandt, *Assessing student learning: New rules, new realities* (pp. 65-92). Arlington, VA: Educational Research Service.
- Rivera, C., Collum, E., Wilner, L. S., & Sia, J. K. (2006). An analysis of state assessment policies addressing the accommodation of English language learners. In C. Rivera, & E. Collum, *A national review of state assessment policy and practice for English language learners* (pp. 1-173). Mahwah, USA: Lawrence Erlbaum.
- Sayer, P. (2013). Translanguaging, TexMex, and bilingual pedagogy: Emergent bilinguals learning through the vernacular. *TESOL quarterly*, 47(1), 63-88.
- Shohamy, E. (2006). Imagined multilingual schools: How come we don't deliver. In O. Garcia, T. Skutnabb-Kangas, & M. E. Torres-Guzman, *Imagining multilingual schools* (pp. 171-183). New York and Tronninge Mose: Multilingual Matters LTD.
- Shohamy, E. (2011). Assessing multilingual competencies: Adopting construct valid assessment policies. *The Modern Language Journal*, 95 (iii), 418-429.
- Sireci, S. G., & Faulkner-Bond, M. (2015). Promoting validity in the assessment of English learners. In *Review of Research in Education*, 39 (pp. 215-252). AERA.
- Sireci, S. G., Li, S., & Scarpati, S. (2003). *The effects of tests accommodations on test performance: A review of the literature*. Commissioned paper by the National Academy of Sciences/National Research Council's Board on Testing and Assessment.
- Sireci, S. G., Scarpati, S. E., & Li, S. (2005). Test accommodations for students with disabilities: An analysis of the interaction hypothesis. *Review of educational research* 75 (4), 457-490.
- Smarter Balanced Assessment Consortium; National Center on Educational Outcomes. (2016). *Smarter Balanced Assessment Consortium: Usability, Accessibility, and Accommodations Guidelines*. Smarter Balanced Assessment Consortium. Retrieved from <http://www.smarterbalanced.org/assessments/accessibility-and-accommodations/>

- So, Y., Zapata-Rivera, D., Cho, Y., Luce, C., & Battistini, L. (2015). Using dialogues to measure English language skills. *Educational Technology & Society*, *18*(2), 21-32.
- Schwabe, F., von Davier, A.A., Chalhoub-Deville, M. (2016). Language and culture in testing. In Leong, F.T., L., Bartram, D., Cheung, F.M., Geisinger, K. F., & Iliescu, D., *The ITC international handbook of testing and assessment*, (pp 300-317), New York, USA: Oxford University Press.
- Thompson, S., Johnstone, C.,J., & Thurlow, M., L. (2002). Universal design applied to large scale assessments (Synthesis Report 44). Minneapolis, MN: University of Minnesota, National Center on Educational Outcomes.
- Thurlow, M. L., & Kopriva, R. J. (2015). Advancing accessibility and accommodations in content assessments for students with disabilities and English learners. *Review of Research in Education*, *39*, 331-369.
- Tollefson, J. W. (2008). Language planning in education. In N. H. Hornberger, *Language policy and political issues in education* (pp. 3-14). New York, NY: Springer Science + Business Media, LLC.
- Trapido, J. (2015). Africa's leaky giant. *New Left Review*, *92* (II). Retrieved from <http://newleftreview.org/II/92/joe-trapido-africa-s-leaky-giant>
- United Nations Educational, Scientific, and Cultural Organisation; International Bureau of Education. (2010). *Donnees mondiales de l'education: Republique democratique du Congo*. United Nations Educational, Scientific, and Cultural Organisation; International Bureau of Education.
- 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.
- Wolf, M., & Leon, S. (2009). An investigation of the language demands in content assessments for English language learners. *Educational Assessment*, *14*(3-4), 139-159.
- Wolf, M., Kim, J., & Kao, J. (2012). The effects of glossary and read-aloud accommodations on English Language Learners' performance on a mathematics assessment. *Applied Measurement in Education*, *25*(4), 347-374.
- Yzquierdo, Z. (1995). History of bilingual assessment and its impact on best practices used today. *New York State Association for Bilingual Education Journal*,*10*, 6-12.
- Zieky, M. (2015). Developing fair tests. In S. Lane, M. Raymond, & T. Haladyna, *Handbook of test development* (2nd ed.,pp. 81-99). New York, USA: Routledge.
- Zumbo, B. (1999). *A handbook on the theory and methods of differential item functioning (DIF): Logistic regression modeling as a unitary framework for binary and likert-*

type (ordinal) item scores. Ottawa, Canada: Directorate of Human Resources Research and Evaluation, Department of National Defense.