Examining PIRLS 2021: Differential item functioning across English and isiZulu language groups

by

SINETHEMBA MTHIMKHULU

Submitted in fulfilment of the requirements of the degree

MAGISTER EDUCATIONIS

in

Assessment and Quality Assurance

FACULTY OF EDUCATION

UNIVERSITY OF PRETORIA

Supervisor: Dr K. Roux

Co-supervisor: Dr M. Mihai

2023

DECLARATION

I declare that the dissertation, which I hereby submit for the degree of M.Ed. Assessment and Quality Assurance in Education and Training at the University of Pretoria is my own work and has not been previously submitted by me for a degree at this or any other tertiary institution.

……

Sinethemba Mthimkhulu

29 September 2023

ETHICAL CLEARANCE

FACULTY OF EDUCATION

Ethics Committee

UNIVERSITEIT VAN PRETORIA
UNIVERSITY OF PRETORIA
YUNIBESITHI YA PRETORIA

RESEARCH ETHICS COMMITTEE

CLEARANCE CERTIFICATE	CLEARANCE NUMBER: **EDU161/22**
DEGREE AND PROJECT	MEd Examining PIRLS 2021: Differential item functioning across English and isizulu language groups
INVESTIGATOR	Mr. Sinethemba Mthimkhulu
DEPARTMENT	Science Mathematics and Technology Education
APPROVAL TO COMMENCE STUDY	30 November 2022
DATE OF CLEARANCE CERTIFICATE	31 August 2023

CHAIRPERSON OF ETHICS COMMITTEE: Prof Funke Omidire

Mr Simon Jiane

Dr Karen Roux

Dr Maryke Mihai

This Ethics Clearance Certificate should be read in conjunction with the Integrated Declaration Form (D08) which specifies details regarding:

- Compliance with approved research protocol,
- No significant changes,
- Informed consent/assent,
- Adverse experience or undue risk,
- Registered title, and
- Data storage requirements.

Fakuiteit Opvoedkunde
Lefapha la Thuto

ACKNOWLEDGEMENTS

As agreed, Mom, here is the dissertation we spoke about. I dedicate this study to my late mother, Mrs B.N Mthimkhulu. You funded my tertiary education right from first year of university till now, I can never thank you enough. I am the man I am today because of your undoubted love, care and inspiration. Every time, when things got tough, I always reminisced about our conversations. Your words *"Funda mntanami"* kept me motivated and going. Please pass my gratitude to God our creator and our ancestors for granting me the courage do this study and the strength to complete it. *Ngiyabonga Mama.*

To my parents, Mr M.J Mthimkhulu, Dad thank you so much, your endless support throughout my academic career, it is held in high regard. Your Sunday calls to check and ask how I am doing kept me going. You always assured me not to worry about my education funds because I had you in my corner, here is what you were funding all along, Dad. What you did for me is beyond the words thank you. To Mom, Mrs N.T Mthimkhulu, thank you for the support, care and encouragement.

To my supervisors, Dr K. Roux, I do not have the words to express my gratitude to you. I am thankful to have compiled this dissertation under your supervision and Dr Mihai's. The academic opportunities you unlocked for me during this study are significant, I am grateful. You never discarded my ideas rather you moulded them into what they are today, a whole dissertation. Thank you for your patience and making certain that I completed this study. Dr M.M. Mihai, I know with you there is no orange, it is either red or green. Your valuable insights have paid off. You did not allow anything less, your supervision and guidance pushed me into doing something great, thank you.

To my siblings, thank you for your continued support and encouragements. To my nieces and nephews, *bantwana bami*, I hope this study can serve as beacon of hope, inspiration and make education fashionable.

Lastly, thank you SinethembaM, you have compiled a Masterpiece.

"Funda mntanami"

PROOF OF EDITING

Cilla Dowse Editing

Dr Cilla Dowse
cilla.dowse@gmail.com
Cell: +27 84 900 7837
(on WhatsApp)

For editing of Master's Dissertations
and
Doctoral Theses

This letter serves to confirm that editing and proofreading was done for:

SINETHEMBA MTHIMKHULU

MAGISTER EDUCATIONIS
Assessment and Quality Assurance
Faculty of Education
University of Pretoria

EXAMINING PIRLS 2021: DIFFERENTIAL ITEM FUNCTION ACROSS ENGLISH AND ISIZULU LANGUAGE GROUPS

Cilla Dowse
23 September 2023

Cilla Dowse
PhD in Assessment and Quality Assurance in Education and Training: University of Pretoria 2014
Basic Editing and Proofreading: McGillivray Linnegar Associates 2008
Programme on Editing Principles and Practices: University of Pretoria 2009
Editing and Proofreading for Academic Purposes: McGillivray Linnegar Associates 2021
Professional Editors' Guild Associate Member, DOW003

ABSTRACT

In today's interconnected and rapidly evolving world, reading literacy stands as a fundamental skill that empowers individuals to thrive intellectually, socially and economically. This study aimed to shed light on the factors that might have a bearing in reading literacy, with a specific focus on isiZulu, using insights derived from the Progress in International Reading Literacy Study (PIRLS) data. The PIRLS trend assessments have revealed that South Africa is experiencing a 'reading crisis' in that all South African Grade 4 learners' achievement falls below the International Benchmark. This study investigated measurement invariance as a possible factor that might have contributed to the difference in learner achievement in isiZulu when compared with English scores. IsiZulu learners came across words such as "i-Hammerhead shark" in the PIRLS achievement booklets that could not be translated into isiZulu by professional translators. It is challenging to expect Grade 4 learners to comprehend and answer items derived from these words if they are difficult to translate or unavailable in isiZulu. To ascertain equivalence, item difficulty was examined to ascertain whether the level of difficulty was the same for the two groups. Functional equivalence was examined to ascertain the behaviour of the items in the two groups. Using Rasch Measurement Theory (RMT), Differential Item Functioning (DIF) was conducted to determine biases towards the learners who completed the PIRLS 2021 assessment in isiZulu. Descriptive statistics revealed that there is a significant difference between the English and isiZulu mean scores. Raw scores and Rasch Analysis displayed some items in the passage lacked metric and functional equivalence.

Keywords: Differential Item Functioning, Equivalence, Measurement Invariance, PIRLS 2021, Reading Literacy, Secondary Analysis, Validity.

LIST OF ACRONYMS

ANA	Annual National Assessment
ANOVA	Analysis of Variance
CAPS	Curriculum Assessment Policy Statement
CEA	Centre for Evaluation and Assessment
COVID-19	Corona Virus Disease 2019
CR	Constructed Response
CPTD	Continuing Professional Teacher Development
DBE	Department of Basic Education
DIF	Different Item Function
DME	Data Management Expert
DMS	Data Management Software
DoE	Department of Education
EGRA	Early Grade Reading Assessments
ETM	Expert Translation Method
FAL	First Additional Language
GDP	Gross Domestic Product
HL	Home Language
ICC	Item Characteristic Curve
IDB-Analyzer	International Database-Analyzer
IEA	International Association for the Evaluation of Educational Achievement
IEB	Independent Examination Board
ILSA	International Large-Scale Assessment
IRG	International Reading Group
IQCM	International Quality Control Monitor
IQMS	Integrated Quality Management System
LiEP	Language in Education Policy
LoLT	Language of Learning and Teaching
MC	Multiple-Choice
NCES	National Centre for Education Statistics
NCS	National Curriculum Statement

NGO	Non-Governmental Organisation
NRC	National Research Coordinator
OBE	Outcomes Based Education
OECD	The Organisation for Economic Co-operation and Development
PanSLAB	Pan South African Language Board
PerSepIdx	Person Separation Index
PIRLS	Progress in International Reading Literacy
PISA	Programme for International Student Assessment
POPI Act	Protection of Personal Information Act
QDG	Questionnaire Development Group
RDG	Reading Development Group
RMT	Rasch Measurement Theory
RSA	Republic of South Africa
RUMM2030	Rasch Unidimensional Measurement Models 2030
SACE	South African Council for Educators
SEACMEQ	Southern and Eastern African Consortium for Monitoring Educational Quality
SASA	South African Schools Acts
SATI	South African Translator's Institute
SD	Standard Deviation
SE	Standard Error
SGB	School Governing Body
SL	Source Language
SPSS	Statistical Package for Social Sciences
TIMSS	Trends in International Mathematics and Science Study
TL	Target Language
UK	United Kingdom
UNC	University of North Carolina
UNESCO	United Nations Educational, Scientific and Cultural Organisations
US	United States
WCG	Western Cape Government

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

CHAPTER 1: INTRODUCTION

1.1 INTRODUCTION AND BACKGROUND

Among other skills, literacy, specifically reading literacy, is a fundamental skill that every South African learner should possess. The South African Minister of the Department of Basic Education (DBE), Angie Motshekga, made it South African education's primary objective to promote reading and ensure that learners attain the basic skills necessary for their schooling career (Motshekga, 2017). The results emerging from the last two rounds of the Progress in International Reading Literacy Study (PIRLS) 2016 and 2021, signalled a "reading crisis" (DBE, 2023; Howie et al., 2017, p.1) in South Africa. Nearly, eight out of ten Grade 4 learners in South Africa cannot read for meaning (Roux et al., 2022), with these learners struggling to answer basic, literal questions and not being able to retrieve explicitly stated information. This finding is concerning, and it needs to be addressed. Spaull (2021) has highlighted that to respect and fulfil the right to basic education, learners should have the ability to read and write with comprehension in their Home Language (HL) at a basic level.

South Africa is situated in a global system where literacy is among one of the fundamental skills that every learner must be in possession of, therefore, it is mandatory and important that every citizen should be able to read for comprehension (Chapter Two Bill of Rights 1996a (29-12), cl 1). Nevertheless, it is also important not to shy away from the fact that the South African education system is still a developing system in the post-apartheid era with it transitioning from a divided education system into a system integrating education for all learners. The South African government allocates large amounts of funds for the development of the education system and according to Mbiza (2018), is spending more money on education than many other countries. For example, in 2016, 15% of South Africa's total budget was spent on education, that is approximately 213.7 billion South African rands. This budgetary allocation indicates that South Africa is intent on developing and upholding its education system. In addition, approximately 18.4% of the country's total Gross Domestic Product (GDP) was spent on the education sector in 2019. With many of

South African schools being classified as Quintile levels 1 and 2[1] (World Bank and the United Nations Educational, Scientific and Cultural Organisations (UNESCO) Institute for Statistics, 2021), the government is working on improving on quality provision education. Despite the large amount spent on basic education to improve school resources, the infrastructure and provide quality education for all, South African performance remains low in International Large-Scale Assessments (ILSA) (*cf.* DBE, 2023; Howie, et al., 2006, 2011, 2017)

South African primary schools have participated in numerous national, regional and international literacy studies to gauge the learners' achievement across different grades (Govender & Hugo, 2020). Ronderos et al. (2021) state that countries participate in the ILSA to gather evidence for policy formation and evaluation, quantify the human capital competitiveness and enhance curriculum and pedagogy in their educational systems. South Africa has participated in various international studies such as the Progress in International Reading Literacy Study (PIRLS), Southern and East African Consortium for Monitoring Educational Quality (SEACMEQ), and the Trends in International Mathematics and Science Study (TIMSS), to gather tangible evidence on the progress made in the basic education areas (Motshekga, 2017). In other words, the country participates in these trend studies to gauge the growth acquired or lack thereof in the education system of the country and inform the improvement of education. For example, Rule and Land (2017) expressed that the current dire state of reading in South Africa is shown by the consistent low ranking in national and international studies. They suggest that teacher training in South Africa should focus on ensuring that a consistent and comprehensive understanding of the teaching of reading is developed to ensure the development of effective pedagogies. The way reading needs to be taught relates to not only developing the skills of decoding, accuracy, pronunciation and oral performance but also comprehension

[1] Quintile refers to the systematic ranking and categorisation of national schools in South Africa. They are ranked from quintile 1 to 5, where 1 is the labelled as the poorest receiving support from the state and 5 as the wealthiest school functioning on their own with little or no support from the government (DBE, 2004).

which would then improve reading levels and have a positive effect on the education system (Rule & Land, 2017).

In large-scale assessments, whether national or international, reading comprehension is vital and if learners have not developed the skill of comprehending the written word, they would be challenged in answering any of the questions, whether the questions are to do with a reading literacy assessment or a mathematics and science assessment.

1.1.2 Measurement Invariance and Learner Reading Literacy

This study, which focused on determining whether possible measurement invariance could have resulted in the difference in the Grade 4 English and isiZulu learners' reading achievement scores during PIRLS 2021, proposes that the possible measurement invariance in the assessment items might have contributed to the difference in the Grade 4 learner achievement. Measurement invariance can be due to a variety of factors including translations. Labuschagne (2015), in investigating the performance of Tshivenda learners in PIRLS 2006, recommended that to ascertain whether translation is a factor in the case of poor performance, different items in the passages in the achievement booklets should be investigated. This study investigated the PIRLS 2021 assessment items which were translated from English to isiZulu.

There are different types of translations, methods and procedures utilised in translation and these were used to translate the PIRLS Grade 4 achievement booklets. Each language has its morphology and orthography which makes the language complex when translation takes place. Mtsatse (2017) states that with African languages, one dialect is recognised as the official language and is thus used in the education curriculum, but there are varying dialects of the same language depending on the context where the language is spoken. The issue of language and dialect emerged during the translation of PIRLS Literacy[2] 2016 passages and items to the indigenous languages which was described as challenging due to the differences in the language

[2] A less difficult version of PIRLS 2016

because of the various dialects (Howie et al., 2017). Affrman (2010) pointed out that translation has been considered one of the main sources of test bias and assessment invalidity.

1.1.2 Validity in Assessments

As part of this study, the validity of one PIRLS passage and its items in the achievement booklet was examined to ascertain both validities, content and construct validity. *The Amazing Octopus* passage which was part of the PIRLS 2021 released instruments in English and isiZulu, was identified for examination. To investigate validity, this study examined metric and functional equivalence to determine the overall validity of the translated instruments. These terms are clarified in the 'Definition of Key Terms' section of this chapter.

The PIRLS assessment instruments are designed and developed in US English, and then participating countries translate and contextually adapt the instruments to the country's national languages (Mtsatse, 2017). Based on the above, the focus of this study was to examine the content (metric equivalence) and construct (functional equivalence) validity of one of the PIRLS 2021 passages after translation from English into isiZulu.

1.1.3 Possible Bias Resulting from Translation Procedures

French (2014) defined test bias as a systematic measurement error where there are parts that are not relevant to the construct measured and they result in higher or lower scores on the measurement of groups examined. In other words, it is when the test has a construct that does not measure the intended outcome and as a result, the test produces varying scores in the groups tested. Test bias is an important consideration in the selection and use of any test; it is, therefore, mandatory that a test must be fair to all the examinees and not biased against one group (Zumbo, 1999). If the assessment is not equally universal to all learners, then a translation error, assessment error or any other type of error could possibly occur. The presence of an error is problematic because the error reduces construct validity, resulting in the test no longer functioning the same across the learners completing the test, placing

some at an advantage of understanding better and others at a disadvantage of not comprehending the test items or the passage.

1.2 PROBLEM STATEMENT

Reading is an action where a piece of text is read for comprehension and interpretation, through understanding and making sense of the words and letters in the text to ultimately form meaning (Tennant, n.d). Mullis and Martin (2021) consider reading literacy as one of the most important abilities learners acquire as they progress in their early school years. It is a foundation for learning across all subjects, as each subject requires an ability to read with understanding. This study was prompted by the poor reading literacy achievement results emerging from the previous round of PIRLS in 2021, which revealed that there is a large discrepancy in achievement scores between the languages in South Africa (DBE, 2023). The main concern of this study is that there is a discrepancy between learners' scores tested in isiZulu compared to those tested in English. This finding is a contemporary issue as the previous cycles, namely, PIRLS 2011 and PIRLS Literacy 2016, found similar discrepancies in the achievement scores between English and isiZulu (Howie et al., 2011; 2017).

The main reason for selecting these languages is that English is one of the two highest-performing languages and isiZulu is the language that is spoken by the majority of South Africans. Roux's (2020) research found that there was no difference in the overall mean scores between English and Afrikaans, however, there was a stark difference when isiZulu was compared to these two languages. For this study, that issue raises questions. English achievement booklets do not need to be translated because they are already designed in English which puts those who took the assessments in English at an advantage. With African languages, some scientific and technological terms are not yet available which makes translation a bit challenging because the African languages have not yet developed and evolved to accommodate modern terminologies. This study closely examined whether measurement invariance could have been a contributing factor to the difference of the achievement scores, especially as measurement invariance could be a result of aspects like translation (Combrinck, 2020).

1.3 RATIONALE OF THE STUDY

The rationale behind this study is that learners in PIRLS 2021, who were tested in isiZulu, achieved a low mean score (267, SE=6.5) which means that they lag behind in terms of formal schooling years when compared to the learners who were tested in English (382, SE=14.5). Note that approximately 40 score points equals to one year of schooling on the PIRLS scale (Howie et al., 2017). Therefore, the question could be asked whether the ILSA test developed in English, adapted to the South African context and translated into the ten remaining South African official languages created a bias in the assessment. With isiZulu, there are different dialects[3] and those do not apply to all the different regions and communities in KwaZulu-Natal or across the country. As much as PIRLS does not test the dialects, they are different, and the test is standardised regardless of the fact that one word for a learner in, for example, the Amajuba district might mean something different for a learner in the Umhlathuze district. The language structures are unique, depending on the region and the district.

This study was essential because it could possibly add value to the ILSA literature, as it was particularly interested in examining the gap between the achievement scores of English and isiZulu learners in the PIRLS 2021 assessment. Secondly, it could prompt scholars to conduct research on the difference in learner performance in the remaining South African languages that are not included in this study. There was a need for this study because it partly forms a continuation of the argument established by Roux (2020), that quality of the translation of the PIRLS instruments may result in item bias or create measurement invariance for other language groups which may be a factor in the difference in the learner achievement. Roux (2020) reported that the PIRLS Literacy 2016 released instruments indicated signs of differential item functioning, but it was not universal to one language. This means that they did not favour any particular language, but the items functioned differently in all the languages investigated. Subsequently, this study suggests that the degree of metric and functional equivalence

[3] The language is not standard or similar to each and every learner taking the test in isiZulu.

found in the assessment, might help explain the difference in the learner achievement in PIRLS 2021. The aforementioned, rationalises the need for this investigation.

1.4 AIM AND OBJECTIVES

The main aim of this research was to determine whether possible measurement invariance could have resulted in the difference in the Grade 4 English and isiZulu learners' reading achievement scores during PIRLS 2021.

Therefore, this study intended to:

- determine the extent to which the Grade 4 English and isiZulu learners' reading achievement scores differ on PIRLS 2021;
- determine the extent to which the difference in achievement can be explained by possible bias (measurement invariance) between English and isiZulu responses during the PIRLS 2021; and
- assess the degree of functional equivalence in the items post-translation.

1.5 RESEARCH QUESTIONS

The study poses the following main research question: *To what extent are The Amazing Octopus items equivalent across English and isiZulu language groups post-translation?*

The following sub-questions were formulated to assist in answering the main research question:

1. How do the overall Grade 4 English and isiZulu learners' reading literacy achievement scores differ on PIRLS 2021?
2. To what extent can the difference in achievement be explained by possible bias (measurement invariance) between English and isiZulu responses during PIRLS 2021?
3. To what degree are the items functionally equivalent post-translation?

1.6 RESEARCH METHODOLOGY OVERVIEW

The purpose of this study was to determine whether possible measurement invariance could have resulted in the difference in the PIRLS 2021 Grade 4 English and isiZulu learners' reading achievement scores. This study was underpinned by the positivism paradigm that requires a scientific approach to research and investigation of the research problem. This paradigm justifies the use of quantitative methods of research that account for numerical data. The study was structured to align with the paradigm because the paradigm allows the use of quantitative approach and methods of research used in this study.

To answer the main question and the set of sub-questions, this study utilised a quantitative approach of gathering and analysing data. PIRLS 2021 quantitative or numerical data were utilised to investigate the current research problem and assisted in answering the main research question and the sub-questions. This approach allowed for the variables to be quantified.

The PIRLS 2021 data was collected by the Centre for Evaluation and Assessment (CEA) at the University of Pretoria in 2021 which means that the current study utilised secondary data. As such, the research design that guided this study is a secondary data research design. To assess reading literacy abilities of learners, the PIRLS 2021 used booklets containing 18 passages that were administered in the field using a group adaptive assessment design. This study identified one passage: *The Amazing Octopus* passage and its items, for the purpose of determining whether possible measurement invariance could have resulted in the difference in the Grade 4 English and isiZulu learners' reading achievement scores during PIRLS 2021.

The main research question of this inquiry focused on determining the extent to which the items in *The Amazing Octopus* passage are equivalent across English and isiZulu language groups. It was necessary for this study to look into the metric and functional equivalence of the items post-translation and adaptations. Therefore, three sub-questions were formulated to assist in answering the main research question. The first sub-question focused on the difference between the Grade 4 learner achievement scores in English and isiZulu. Thus, this study relied on descriptive statistics such as

raw scores, raw mean scores and percent correct to answer to sub-question 1. To ascertain whether possible measurement invariance can help explain the difference in the achievement scores, inferential statistics such as ANOVA statistics (Analysis of Variance) were utilised to answer sub-question two. Lastly, sub-question three was formulated to observe item behaviour in the different languages investigated. Item Characteristic Curve (ICC) graphs were used to observe the functionality of the item.

Since this study made use of PIRLS 2021 data, it adopted the Stratified Two-Stage Cluster sampling procedure used by the PIRLS 2021 cycle to sample the participants in their study. Due to the complexities of the PIRLS 2021 data, two specific software programmes were used for this study, namely, International Database-Analyzer (IDB-Analyzer) and Statistical Package for Social Sciences (SPSS). The former was specially designed to analyse PIRLS data as it takes into consideration the weights of the data. In addition to the software used, Rasch Unidimensional Measurement Models 2030 (RUMM2030) was used to test for Different Item Function (DIF) and observe item behaviour through statistical outputs produced by the software in the chosen items in *The Amazing Octopus* passage.

1.7 DEFINITION OF KEY TERMS

Reading literacy is defined by Mullis and Martin (2021, p. 6) as

> The ability to understand and use those written language forms required by society and/or valued by the individual. Readers can construct meaning from texts in a variety of forms. They read to learn, to participate in communities of readers in school and everyday life, and for enjoyment.

In other words, reading literacy is active engagement with various literary and informational texts to gain an understanding of what is discussed in the passages.

Measurement invariance is seen as a measurement instrument functioning the same way across varied conditions (Millsap, 2011). The implication is that the same instrument should measure the same construct across different groups. It is further articulated by Putnick and Bornstein (2016:71) that measurement invariance

"assesses the psychometric equivalence of a construct across different groups". Therefore, measurement invariance is concerned with whether the instrument is interpreted the same way by varying groups responding to the instrument. For instance, if the English and isiZulu respondents interpret the items differently, then there might be evidence of the inequivalence in the instrument.

Validity refers to the assessment assessing what it intends to in the first place (Chen, 2010). Validity has to do with the designed assessment serving its purpose of assessing the predetermined assessment outcomes. In assessment, validity is crucial because the assessment data is used to make decisions about the learners' learning and determine their abilities in what is tested.

In the arena of assessments, *equivalence* is seen as the sameness of the translated instruments from one language to the other so that the scores can be compared across the languages (Aegisdóttir et al., 2008). Equivalence is concerned with the assessment instruments being identical in two or more languages after the translation processes have taken place. The assessment instrument must be similar in all the languages in which the assessment is to be translated and administered.

1.8 STRUCTURE OF THE DISSERTATION

Chapter 1: This first chapter referred to PIRLS trend cycles and particularly the South African findings, which highlighted a concern about the reading literacy of learners identifying it as a 'reading crisis'. The Minister of Education has made it the nation's aim to improve reading literacy among South African children. This first chapter outlined the problem and included the rationale for undertaking this study. The objectives of this study are outlined in this chapter to situate the study in literature and to articulate its value and possible contribution to the ILSA. This chapter provided a layout of the research aims and research questions investigated in this study. It also included the definition of key terms that are predominant in this study.

Chapter 2: This chapter provides an extensive overview of the PIRLS 2021 study as the main source of this research. When discussing the PIRLS 2021 background and history, reference is made to its previous cycles to establish the connection between

the studies and to demonstrate the development of the PIRLS study over the years since it was introduced in 2001. Equally important, is the PIRLS assessment framework that this chapter discusses followed by the PIRLS instruments (questionnaires and achievement booklets). This chapter also highlights the assessment design PIRLS 2021 used in the field to administer the achievement booklets. Translation procedures and translation and adaptations of the PIRLS 2021 achievement booklets are discussed to establish the process followed by PIRLS in their translation of the instruments. The data collection processes and analysis of PIRLS data are discussed.

Chapter 3: This chapter reviews the literature, discussing the main themes identified for this study. The themes encompass reading literacy as a fundamental skill, translation complexities that can result in possible measurement invariance, equivalences in assessments and the concept of validity in assessments. Equivalence is described in this chapter and metric and functional equivalence are distinguished in great detail. Issues concerning validity post-translation are reviewed. Regarding validities, different kinds of validity (content and construct validity) are discussed and elaborated on. The theoretical framework which underpins the study forms part of this chapter. This study is viewed and undertaken under the theoretical framework guide of the curriculum process framework, which makes room for the achieved curriculum (PIRLS 2021 achievement scores) to be studied while taking into account the intended and the implemented curriculum. It also considers the context and the antecedents as notable elements when dealing with the curriculum side of the framework.

Chapter 4: This chapter is concerned with the research design and the methodology of the study. Since the study draws from the PIRLS 2021 data, this study takes the form of a secondary analysis. As such, this chapter reports on the original study's sampling and data collection methods and procedures. The current study made use of a secondary research design as it allows the use of quantitative methods of research to investigate the research problem presented in this study. The reason for this approach is that this inquiry sought to determine whether the items in *The Amazing Octopus* passage are equivalent across English and isiZulu language groups. The data are analysed through the use of IDB-Analyzer in conjunction with SPSS.

Additionally, RUMM 2030 is used to look into DIF through the Rasch Measurement Theory (RMT). This chapter also provides a brief overview of the statistics to be used. The discussion of methodological norms forms part of this chapter.

Chapter 5: This chapter presents the research findings. The main aim of this study focused on determining the extent to which the items of *The Amazing Octopus's* passage are equivalent. This study conducted descriptive statistics that were used to address sub-question 1. Based on the findings of sub-question 1, it was then necessary for this study to further explore the data by conducting inferential statistics that included Rasch Analysis that displayed which of the PIRLS 2021 items showed differential item functioning thus lacking metric and functional equivalence. Through the ANOVA Statistics output produced by RUMM2030, this study graphically illustrates which of the items function differently and exactly[4] how they function differently for the two language groups investigated. Using the overall data findings, this study is able to determine the extent of the equivalence of *The Amazing Octopus* passage items.

Chapter 6: This chapter presents the summary of the research findings per sub-question. Based on the findings in the previous chapter, this study is able to reach systematically generated conclusions that are able to answer the research's main question and sub-questions. Thereafter, the strengths and limitations of this study are discussed in this chapter. After a careful consideration of the results and findings, this study makes recommendations for policy, practice and further research. The potential contribution of this inquiry is discussed before concluding the study with final thoughts.

[4] Using Item Characteristics Curve (ICC) graph it can be determined that in which class-interval (logits scale) and probability scale did the item present DIF.

CHAPTER 2: PROGRESS IN INTERNATIONAL READING LITERACY STUDY

2.1 INTRODUCTION

The Progress in International Reading Literacy Study **(**PIRLS) is an international comparative study that assesses Grade 4 learners' reading literacy across more than 50 countries. Johansone (2023) postulated that to ensure a consistent and uniform approach to attain high quality and internationally comparable data, all participating countries are expected to follow a set of standardised operations and procedures. As the current study draws from PIRLS cycles, the PIRLS 2021 assessment is discussed in this chapter focusing on the standardised operations and procedures utilised by participating countries including South Africa.

An overview of the history and origins of PIRLS is described in Section 2.2 followed by the PIRLS reading framework (Section 2.3). Next, the PIRLS contextual questionnaires (Section 2.4) and assessment booklet design (Section 2.5) are discussed. Section 2.6 deliberates on PIRLS translations and adaptations of instruments, followed by a discussion on the PIRLS sampling design (Section 2.7). Following that, an overview is provided of the PIRLS main study data collection (Section 2.8). Data capturing, processing and analysis are reviewed in Section 2.9. Thereafter, the quality assurance procedures used by PIRLS is described (Section 2.10). Lastly, a conclusion is provided.

2.2 HISTORY AND ORIGINS

The International Association for the Evaluation of Educational Achievement (IEA) is a Non-Governmental Organisation (NGO) that has undertaken several ILSAs such as PIRLS and TIMSS. IEA PIRLS is aimed at measuring the reading literacy abilities of Grade 4 learners in participating countries across the world (Mullis & Martin, 2021) and assessing international trends in reading comprehension of Grade 4 learners. It is expected that at this level of schooling, learners demonstrate reading abilities and are reading to learn (Von Davier et al., 2023; Mullis et al., 2017). PIRLS intends to provide the participating countries with reading comprehension trends across different

cycles, including the learners' educational opportunities through the gathering of data using assessment booklets and background questionnaires. PIRLS views reading as an interactive process between the text and the reader of the text (Mullis & Martin, 2015). PIRLS 2021 was the fifth round of reading assessments; however, for South Africa it was the fourth round of taking part in the PIRLS assessments (DBE, 2023).

Internationally, PIRLS is conducted by the IEA at Boston College, United States and works closely with IEA Hamburg and IEA Amsterdam. In South Africa, it is handled by the CEA located at the University of Pretoria (DBE, 2023; Howie et al., 2017). PIRLS takes place in five-year cycles, and it was first introduced in South Africa in 2006. The proceeding of PIRLS assessments is endorsed by the Department of Basic Education (DBE) in South Africa.

Currently, PIRLS has two assessments namely, PIRLS and digitalPIRLS. The PIRLS study is made up of a state-of-the-art reading assessment that measures the international trends in reading achievement. These assessments come with an extensive set of questionnaires used to gather the learners' background data that include contexts for learning and educational opportunities to learn (Mullis & Martin, 2021). The PIRLS achievement booklets consist of different passages and items to test learners' reading literacy using Multiple-Choice (MC) and open/Constructed-Response (CR) questions.

After each round of PIRLS, a summary report of the findings is released (DBE, 2023; Howie et al., 2006; 2011; 2017) presenting the results and findings of the data that researchers have gathered in a particular cycle. Pre-PIRLS 2011 and PIRLS Literacy 2016 reported that an average South African Grade 4 learner cannot read for meaning, that is because their average score fell below the international set mean. This means that the South African Grade 4 learner performed below the international average of 500 score points. In PIRLS 2021, South Africa achieved 288 score points with a Standard Error (SE) of 4.4 (DBE, 2023). Overall, South Africa learners have been performing poorly across the PIRLS trends and have been unable to achieve the low international benchmark of 400 score points. This concerning finding by PIRLS international trend results called for national awareness that emphasises the importance of reading, making it a national priority in South Africa.

This present study is based the PIRLS 2021 cycle. Table 2.1 depicts the overall mean scores and the SE for each language assessed during PIRLS 2021.

Table 2.1: PIRLS 2021 scores across all South African languages.

Languages	PIRLS Literacy 2016
English*	382 (14.5)
Afrikaans	387 (15.4)
Sesotho	258 (10.9)
IsiNdebele	255 (9.8)
SiSwati	257 (6.2)
IsiZulu*	267 (6.5)
Xitsonga	223 (13.2)
Tshivenda	255 (9.0)
Setswana	211 (7.3)
IsiXhosa	254 (8.1)
Sepedi	216 (8.7)

(Source: Progress in International Reading Literacy Study 2021 South African Preliminary Highlights Report (p.4), by DBE, 2023, Department of Basic Education. Copyright 2017 by Department of Basic Education)

According to the Minister of Education, English is one of the highest-performing languages in PIRLS, alongside Afrikaans (Motshekga, 2017; 2023) which was the highest performing language in PIRLS 2021 with 387 (SE=15.4[5]) points. In contrast, all South African languages performed poorly with the lowest performing language being Setswana with 211 (SE=7.3). However, all languages performed well below the low international benchmark (400-474). In PIRLS studies, African languages (specifically African languages spoken in South Africa) learners are achieving way below Afrikaans and English learners. Of concern for this study, is that isiZulu learners underperform in comparison to learners who took the test in English (DBE, 2023);

[5] When interpreting the PIRLS results, it is vital to note the SEs.

therefore, these languages became the focus of the study, as indicated with an asterisk in Table 2.1.

2.3 PIRLS READING FRAMEWORK

The change from *learning to read* to *reading to learn* is considered a crucial stage in learners' development as readers, therefore, PIRLS focuses on the achievement and reading experiences of learners in Grade 4 across the participating countries. Mullis and Martin (2021) acknowledge that the reading purpose functions are integrated with the comprehension processes, meaning that they work interdependently. PIRLS tests the *purpose of reading* and the *processes of comprehension* with achievement booklets while reading behaviour and attitudes are gathered using background questionnaires (Mullis & Martin, 2021).

2.3.1 Purposes for Reading

Reading at a young age requires learners to read for different purposes like reading for pleasure, learning and acquiring information. The PIRLS passages are directed towards enhancing and assessing those abilities; hence the PIRLS passages contain equal proportions of material that are directed toward assessing each purpose. The achievement booklets are guided by a matrix of two reading purposes (a) *literary* and (b) *informational.* The PIRLS passages are classified by their primary purpose and the accompanying items that are aimed at addressing the purpose of reading for that specific passage. PIRLS *literary* passages have items assessing aspects like theme, plot, characters and setting whereas those that are classified as *informational* are paired with items that are asking what the passage is about (Mullis & Martin, 2021). Reading for literary experience is regularly matched with fictional material while reading to acquire and use information is most likely to be linked with informative texts and instructional materials (Roux, 2015). PIRLS passages are multimodal because they include both written texts and pictures that communicate meaning (Mullis & Martin, 2017) and enhance the written word.

The *literary* reading passages allow the readers to interact with the text to comprehend the events, setting, characters and atmosphere associated with the passage. This

requires the readers to bring their own experiences, feelings and appreciation of literature (Mullis & Martin, 2021). The *literary* experience passages allow the learners to experience situations that they have not yet encountered. The passages may be structured in a form of a narration or dialogue between the characters, which requires the learner to follow the sequence of the passages chronologically and logically. These passages take the form of fiction, which implies that the readers, who in this case are Grade 4 learners, need to activate their creativity to imagine what is happening in the passages. Sometimes pictures are included to aid learners' understanding of the passage.

The primary purpose of *informational* texts is to disseminate information. PIRLS informational texts range from scientific, historical, geographical and social information. The structure and format of the contents of the informational passages differ according to the type of informational passages. The informational texts utilised by PIRLS assessments reflect learners' authentic experiences of reading informational texts both in and out of school. PIRLS informational texts are often compiled by authors that are familiar with a young audience of readers (Mullis & Martin, 2021). Experts from all participating countries, who are familiar with the Grade 4 level reading materials, compile the booklets. In contrast to reading for a literary experience where learners read for imagination, informational texts aid the learner in understanding how the world functions as reading to acquire and use information introduces the learners to information about the world and taps into the learners' reasoning skills (Roux, 2015).

2.3.2 Processes of Comprehension

There are multiple forms of creating meaning from the text and varying reading situations call for learners to create meaning in varying ways (Mullis & Martin, 2021). Some learners might make meaning of what they are reading by focusing on and retrieving clearly stated information while others may make straightforward inferences. Some learners may interpret and integrate ideas and information to comprehend the notions presented in the piece of information they are reading. In contrast, some may examine and evaluate the content, language and textual elements of the passage. All these aspects occur to comprehend what learners are reading in order to answer the set of items following the piece of text. As a result, Mullis and Martin (2021) included

the classification of the *processes of comprehension* used by PIRLS in the PIRLS 2021 assessment framework. It is crucial to note that one learner may use more than one process of comprehension to make sense of what they are reading. Table 2.2 shows the PIRLS 2021 processes of comprehension and the percentage of items allocated toward the coverage of the comprehension process.

Table 2.2: PIRLS Processes of comprehension

Literary Experience (50%)	Different categories	Acquire and use information (50%)
20%	Focus on and retrieve explicitly stated Information	20%
30%	Make straightforward inferences	30%
30%	Interpret and integrate ideas and information	30%
20%	Evaluate and critique content and textual elements	20%

(Source: PIRLS, 2021 Assessment Framework (p. 8.), by Mullis and Martin, 2021, TIMSS & PIRLS International Study Center. Copyright 2019 by TIMSS & PIRLS international Study Center)

The PIRLS' processes of comprehension are guided by meta-cognitive processes and strategies. Reading strategies come in two categories, cognitive and meta-cognitive. Cognitive strategies deal with the learner making use of their prior knowledge, rereading and changing the reading speed for comprehension purposes, whereas, meta-cognitive is the self-monitoring and self-regulation of the cognitive strategies (Howie et al., 2017). The PIRLS reading comprehension processes can be used as guidelines for the teaching of reading literacy in the Foundation Phase. One of the advantages of PIRLS being conducted in South Africa is that the previous cycles of PIRLS have alerted the DBE that South African young readers need more reading materials (Howie et al., 2017). Furthermore, the advantage of findings emerging from PIRLS data is that a number of studies have identified factors that might have an association with learner reading literacy development (Labuschagne, 2015; Mtsatse, 2015; Roux; 2015; Roux 2020); thus, evidence indicates that necessary changes are needed in order to improve reading literacy. Teacher qualifications, teachers'

pedagogical knowledge, resources. learner attitudes and parental involvement have been investigated as factors that might have an association with reading literacy.

Each PIRLS 2021 achievement booklet includes a set of items specifically designed to test each process of comprehension. The various processes of comprehension permit the learners to make meaning in various forms and adjust their approaches to reading for comprehension. Equally important is prior knowledge and experiences that the learners bring to the classroom to help them understand what they are reading. The four processes of comprehension have designated items that range from easy to difficult, inferring that the nature[6] of the text influences the complexity of the items across and within the processes of comprehension. Each of the processes of comprehension are discussed in detail in the next sub-sections indicating how they fit into the PIRLS 2021 assessment framework.

Focus on and retrieve explicitly stated information: The readers of any text focus their attention to different parts of the clearly stated information (Mullis & Martin, 2021). Since their attention is located on different parts of the text, they also use different ways to focus on and retrieve the different parts of information communicated in the text. Normally, this process of comprehension entails that the learner focuses on the word, phrase and/or sentence level to make meaning of what they are reading (Mullis & Martin, 2015). The focus of learners may be on the information that indicates the theme or the setting of the narrative in the passages, or rather on the information that clearly describes the character in the text. Following the above, learners may be asked to extract information from the text that is plainly stated in the text to respond to certain items. This is to gauge their developing understanding of some aspects of the passage meaning and to answer the questions they may have about the text. This process requires the reader to recognise the relevance of the extracted idea/s to the information sought.

[6] The nature of the text has to do with the length, syntactic complexity, the abstractness of ideas, organisational structure and cognitive demand of the texts.

Reading tasks that are connected with this kind of processing include:

- Identifying information that is relevant to the specific goal of reading;
- Looking for specific ideas;
- Searching for definitions of words or phrases;
- Identifying the setting of a story (e.g., time and place); and
- Finding the topic sentence or main idea (when explicitly stated) (Mullis & Martin, 2021, p.13).

Make straightforward inferences: As learners navigate through the texts, they try to create different meanings by making inferences regarding the ideas and the information that is not clearly stated (Mullis & Martin, 2021), which allows them to transition past the surface of filling in gaps in meaning. As much as some of the conclusions may be straightforward, based on the information visible in the text, learners need to make connections with more ideas for comprehension. Making connections with other pieces of information in the text depends on the readers' abilities because some may be quicker to make the links while others may take time (Mullis & Martin, 2015). Roux (2015) postulated that this type of processing depends on more than just sentences at the immediate phrase level. Theoretically, it may require the learner to infer what the phrase or the word means beyond the literary meaning. For Mullis and Martin (2021) that means that the ideas may be clearly stated but the connection between them is not, which necessitates that the learners derive the meaning for themselves. The items and the correct response in this process may be paraphrased and/or extracted exactly from the text, and thus classified as straightforward inferences (Mullis & Martin, 2021).

The following includes the reading tasks that may demonstrate this kind of processing:

- Inferring that one event caused another event;
- Giving the reasons for the characters' actions;
- Describing the relationship between the characters; and
- Identifying which section of the text would help for a particular purpose (Mullis & Martin, 2021, p. 14).

Interpret and integrate ideas and information: In this comprehension process, the readers focus on local and global meanings, or they may relate details to the overall themes and ideas of the text (Mullis & Martin, 2021). As the learners engage in the process of interpreting and integrating ideas and information, they attempt to develop a more specific and complete understanding of the passage (Mullis & Martin, 2015). This is done by integrating their personal knowledge and experiences with the meaning within the passage. The interpretive process requires the reader to make connections that are implicit (not clearly expressed) and open to interpretation informed by their perspective (Mullis & Martin, 2021). The learners integrate their experiences to help interpret the ideas and make connections within the texts to obtain the overall meaning of the passage.

The reading tasks in this process of comprehension include:

- Discerning the overall message or theme of a text;
- Considering an alternative to the actions of characters;
- Comparing and contrasting information;
- Inferring a story's mood or tone;
- Interpreting a real-world application of text information; and
- Comparing and contrasting information presented within and across texts (Mullis & Martin, 2021, p.15).

Evaluate and critique content and textual elements: For this comprehension process, the focus is on critically considering the text itself, through the reader stepping back, evaluating and critiquing the text (Mullis & Martin, 2021). Here the readers must draw on their interpretation and comprehension of the passage against their knowledge of the world (Roux, 2015). The learners' knowledge of language usage, text structure and presentational features such as pictures, tables and illustrations are activated. For the learners to critically evaluate and critique the text, they make use of their knowledge. The items in this process of comprehension are categorised as evaluating and critiquing because they prompt the learner to evaluate and critique the text read. If the acceptable response includes a judgement and/or evidence about the aspects

asked, the learner could either argue one side of the chosen judgement or both and thereafter present evidence to support the claim made (Mullis & Martin, 2021).

This process of comprehension includes the following reading tasks:

- Judging the completeness or clarity of information in the text;
- Evaluating the likelihood that the events described could happen;
- Evaluating how likely an author's argument would be to change what people think and do;
- Judging how well the title of the text reflects the main theme;
- Describing the effect of language features, such as metaphors or tone;
- Describing the effect of the graphic elements in the text or website;
- Determining the point of view or bias of the text or website; and
- Determining an author's perspective on the central topic (Mullis & Martin, 2021, p.16).

2.4 PIRLS 2021 CONTEXTUAL QUESTIONNAIRES

In addition to the achievement booklets completed by learners to assess their reading ability, PIRLS uses contextual questionnaires to collect data about the school, home and learner factors that have a role in the development of reading literacy (Johansone, 2023). In February 2018, to prepare for PIRLS 2021, the Questionnaire Development Group (QDG) and the National Research Coordinators (NRCs) of each participating country met to refine the questionnaires, deleting topics that were no longer useful and including new relevant topics and refining scales (Martin & Lui, 2019). The PIRLS questionnaires include home, teacher, school, learner and a separate questionnaire completed by the NRC of every participating country (Johansone, 2023). These contextual questionnaires, discussed below, are linked to different contexts that play a role in learner reading literacy achievement.

- The *home questionnaire* is linked to the home context directed to the parents and/or caregivers of the learners taking part in PIRLS. It investigates aspects like languages spoken at home, parents' reading activities, their attitudes towards reading and their education levels. Additionally, it also looks into

early childhood education, which includes literacy and numeracy (Mullis & Martin, 2021).

- The *teacher questionnaire* investigates the classroom contexts for reading instruction, it also gathers information about the teacher's characteristics (Mullis & Martin, 2021). It is completed by teachers of the participating learners.
- The *school questionnaire* is completed by principals of the participating schools and is aimed at gathering information about the school environment, school resources and technology in the school (Mullis & Martin, 2021).
- The *learner questionnaire,* completed by the learners sampled for the assessment, is directed towards the collection of data about the learners' home environment, their schooling experiences and attitudes toward reading (Mullis & Martin, 2021). PIRLS reading behaviour and attitudes incorporate learner self-concept, learner motivation, readiness to learn and learner reading literacy behaviour (Mullis & Martin, 2015). All these components contribute toward the learners' achievement in literacy.
- Lastly, the *NRC questionnaire*, which gauges the education system of that participating country, examines aspects such as the curricula and pedagogy used in that country.

Mullis and Martin (2021) explain that the purpose of the gathering of reading behaviour and attitude data, including the other data collected with the questionnaires, is to help interpret learner achievement scores. However, for the purpose of this research, the questionnaires are not included as part of the instruments to be used.

2.5 PIRLS ASSESSMENT BOOKLET DESIGN

PIRLS 2021 employs a group-adaptive assessment design, which means that all participating countries administer the same reading passages and items, but the rate at which different test forms are distributed and tailored to the population differs by country (Table 2.3) (Wry & Mullis, 2023). This innovative adaptive design improves PIRLS measurement of reading at all levels of the distribution for countries with varying reading proficiency while also increasing learner engagement (Mullis & Martin, 2021).

It also allows the participating countries to tailor the international instruments based on their Grade 4 learner reading proficiencies. The design is adaptive based on each education system's context while maintaining the international standards of comparable assessments.

For South Africa, the group adaptive design means that 70% of the passages administered are less difficult and 30% of the passages more difficult. During the administration of the assessment, learners seated next to each other do not get the same achievement booklets, which means that learners who seated next to each other answer different booklets (DBE, 2023).

Table 2.3: Table illustrates the PIRLS 2021 Group adaptive assessment design

Reading Purpose	Difficulty Level	Passage Label*	Passage Name*
Literary	Difficulty	LitD1 (06)	Shiny Straw (06)
		LitD2 (16)	Oliver and the Griffin (16)
		LitD3 (21)	New LitD3 passage (21)
	Medium	LitM1 (16)	Pemba Sherpa (16)
		LitM2 (21)	New LitM2 Passage (21)
		LitM3 (11)	The Empty Pot (11)
	Easy	LitE1 (21)	New LitE1 Passage (21)
		LitE2 (11)	The summer my Father Was 10 (11)
		LitE3 (16)	Library Mouse (16)
Informational	Difficult	InfD1 (11)	Where's the Honey? (11)
		InfD2 (16)	Icelandic Horses (16)
		InfD3 (21)	New InfD3 passage (21)
	Medium	InfM1 (16)	How Did We Learn to Fly (16)
		InfM2 (21)	New InfD3 Passage (21)
		InfM3 (06)	Sharks (06)
	Easy	InfE1 (21)	New InfE1 Passage (21)
		InfE2 (11)	Training a Deaf Polar Bear (11)
		InfE3 (16)	Hungry Plant (16)

**The number in parentheses is the assessment year in which the passage was first used*

(Source: PIRLS 2021 Assessment Framework (p.61), by Mullis et al., 2021, TIMSS & PIRLS International Study Center. Copyright 2019 by TIMSS & PIRLS International Study Center)

The PIRLS assessment booklet design is structured in a way that some of the passages and items from the previous round are carried over to the new implemented round to measure changes in the learner's achievement across different cycles (Ebbs & Wry, 2017). The carried-over items and passages are referred to as trend passages and achievement items. The above denotes that some of the passages that were administered in PIRLS Literacy 2016 were also used in PIRLS 2021 unaltered. If any changes are made, they are documented and sent during the verification process of translation. PIRLS uses a collaborative process to develop test items and scoring guides for field testing and the main study, each country taking part in the trend also participates in the development and reviewing of the items. Post-field testing, the TIMSS and PIRLS International Study Center analyses the quality of the items and thereafter decides which ones remain and which items are eliminated. The analysis provides a summary of item statistics for each item that was tested in the field (Wry & Mullis, 2023).

PIRLS adapts its assessment framework each cycle using the information provided by the participating countries' NRC and descriptions of the curriculum and instruction discussed in the PIRLS 2021 encyclopaedia. For instance, PIRLS Literacy 2016 used the rotation test design and transitioned in 2021 to utilising the group-adaptive design. That is done to keep their assessment framework in line with the current curricula and standards of the participating countries, keeping their ideas contemporary and educationally relevant (Mullis & Martin, 2021). The PIRLS 2021 assessment design maximises the data gathered from the achievement booklets while minimally adapting the existing procedures and time requirements (Wry & Mullis, 2023).

2.6 PIRLS TRANSLATIONS AND ADAPTATIONS OF INSTRUMENTS

In assessments, translation refers to the creation of different language versions of an assessment that are linguistically equivalent in both the Source Languages (SL) and Target Language (TL) (Ronderos et al., 2021). In the context of PIRLS, it entails converting the achievement booklets to the national languages of the participating

countries. Therefore, in South Africa, the Grade 4 learners were tested using the Language of Learning and Teaching (LoLT) used at the Foundation Phase level (Roux et al., 2022). The IEA as an ILSA administrator, designs and develops the international instruments applicable to all countries. Thereafter, they disseminate the instruments to the participating countries to be translated using a procedural manual sent with the instrument package (Johansone, 2023). Each participating country is responsible for the translations of the instruments originally designed in United States (US) English to their national languages.

To ensure that the translation process is of high quality and internationally comparable across all the participating countries, countries are required to follow the standard international agreed-upon procedures during the process of converting and intentionally altering the instruments to fit the context of the country (Ebbs & Wry, 2017). It is noteworthy to mention that since PIRLS minimally changes their operations and procedures (Johansone, 2023), the preparation of PIRLS 2021 national instruments was relatively similar to the procedures used in the PIRLS Literacy 2016 trend cycle (Ebbs et al., 2023).

In South Africa, the contextual questionnaires and assessment instruments were translated into the remaining ten official languages. English was excluded because initially the questionnaires and assessments instruments were designed in the English, specifically US English was used (DBE, 2023; Howie et al., 2017). However, it is crucial to mention that in South Africa, the English assessments were adapted to United Kingdom (UK) English. For the South African study, the translation procedures included translations, back translation, translation verification, formatting, and layout verification (Roux, 2020), before the printing of the instruments for implementation.

Translations were developed through an extensive collaborative engagement and process between a team of translation experts and a team of reviewers with extensive knowledge of the terminology used in the participating countries (Mullis et al., 2017). The procedure of translation and adaptation was done through an interactive process involving the International Reading Group (IRG) consisting of specialists from across the world, the NRC and the language professionals employed by the NRC (Labuschagne, 2015). The aim was to create equivalent passages for all learners and

to ensure that all facets of the assessment translations were accurate across the different countries taking part. However, creating equivalent versions of the English assessment to the other African languages can prove challenging because phrases and vocabulary may not be available in those languages (Howie et al., 2017). After the international version of the instruments is sent to the participating countries, each country begins with their translation and adaptation of the international instruments into the national versions. The NRC of each country is tasked with overseeing the process of instrument preparation that includes translations and adaptations (Johansone, 2023).

The National PIRLS Report (DBE, 2023) explains that participating countries hired translators. To increase translation effectiveness each participating country were required to hire professional translators and reviewers with the following requirements:

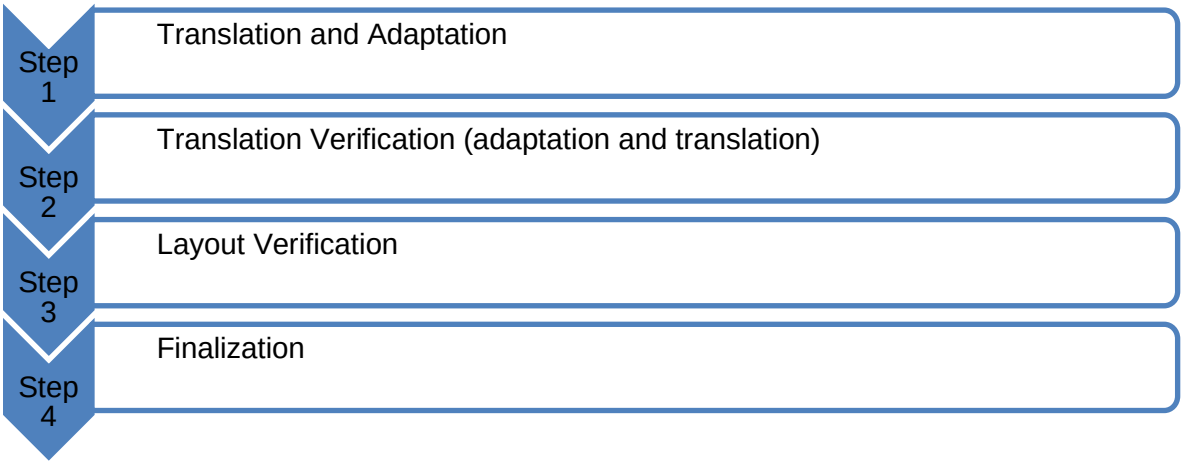
- Excellent knowledge of English;
- Excellent knowledge of the target language (TL);
- Understanding of the country's cultural background; and
- Experience in translating texts at the target Grade(s) (Ebbs et al., 2023, p. 13).

In South Africa, translators registered with the South African Translators' Institute (SATI) were hired to translate and review the translation of the instruments and to pay attention to the readability of the passages in the TL.

2.6.1 Translations and Adaptation Procedures

PIRLS instruments undergo an intensive process of translation to try and validate the authenticity of the instruments and to ensure reliability in the instruments before they are tested in the field. This process makes the data yielded comparable (*cf.* Figure 2.1). The process of intensive translation of the assessment booklets takes place with the intention of creating equivalent passages for all learners while maintaining the quality of the instruments set by the IEA (Howie et al., 2017). Each participating country is responsible for the translations and adaptations of the instruments originally

designed in US English to their national languages. This process is depicted in Figure 2.1.



(Source: Korsnakova et al., 2020, p. 98)

Figure 2.1: The translations-related steps in PIRLS

In South Africa, the PIRLS assessment instruments were translated into ten official languages. To ascertain whether measurement invariance is a factor in the variation of the achievement scores in the languages identified post translation, item bias was closely examined through the use of DIF.

According to the quality assurance programme of translating and adapting instruments, after the first translation and adaptation process, the instruments must undergo a formal external review of the translated and adapted national versions (Ebbs & Wry, 2017). Labuschagne (2015) asserted that the PIRLS translation procedures stipulate that after translation to the national languages, the instruments must be sent to the NRC for verification and to the TIMSS and PIRLS International Study Center before they are used or sent into the field. PIRLS requires multiple rounds of reviews by both linguistic and assessment experts to check that the international version is equivalently translated into the national languages of the participating countries. The reviews seek high quality in the translation processes and allow the adaptations of instruments to each country's context and education system while keeping in mind that the instruments must be comparable across the participating countries. This means that participating countries translate and adapt

their instruments to their language/s of instruction according to certain guidelines to maintain the comparability of the assessments.

The reviews take place in a two-stage format, translation verification and layout verification. Translation verification is the responsibility of IEA Amsterdam and layout verification is handled by the TIMSS and PIRLS International Study Center (Ebbs et al., 2023). The participating countries interact with different handlers at different stages of the translation process to ensure the quality of instruments. The two-stage format of formal review takes place twice, once before field testing and again before the main study. During field testing, comprehensive notes are compiled by the experts and describe any translation and adaptation issues that might require more adaptations of the instrument specifically for the country.

During translation, PIRLS makes use of back-translation. Behr (2017) describes back translation as the re-translation of a translated instrument back to the language of origin and the consequent comparison of the original version and the back-translated version of the instruments. The process of back translation necessitates that the translated version of the instrument be translated back to the SL and thereafter compare the two versions. This procedure takes place with the intention of establishing a linguistic equivalence in the translations (Roux et al., 2022).

The TIMSS and PIRLS International Study Center relies on the professional evaluation of the specific country's NRC and the experts hired specifically by the country. IEA Amsterdam is the organisation responsible for all participating countries' verification of translations and works in collaboration with the TIMSS and PIRLS International Study Center to manage an extensive series of verification checks to ensure that the PIRLS passages, items and other instruments are comparable across the participating countries (Mullis & Martin, 2021). During translation verification, the verifiers responsible compare the nationally translated text to the international versions and then give detailed feedback to improve the accuracy and comparability of the nationally translated texts. The process of translation and translation verification also occurs in a two-stage format, firstly during the field testing and secondly, before the instruments are used in the field for main data collection, similar to the reviewing process (Johansone, 2023). The two-stage format review in translation and translation

verification is also referred to as the Expert Translation Method (ETM). Post-completion of the translation verification, the nationally translated instruments are then sent to the TIMSS and PIRLS International Study Center for layout verification (Ebbs & Wry, 2017; Ebbs et al., 2023; Johansone, 2023).

Formatting and verification of layout is seen as a process of comparing the national instruments to the international version of the instruments in terms of pagination, text formats, order of items, location of graphics and response options (Korsnakova et al., 2020). Any noted discrepancies are sent back to the specific country before printing begins. At this stage of translation, the verifiers identify any errors to ensure that the national instruments follow the international format in the procedural manual guide. Additionally, this stage also verifies whether the national adaptations made to the instruments do not deviate from the international standards that may possibly unduly influence international comparability of the assessments (Ebbs & Wry, 2017; Johansone, 2023).

The way in which the texts are organised and formatted differs. They differ to a certain degree that ranges from the sequential ordering of written text to snippets of words and phrases arranged with pictorial and tabular data (Mullis & Martin, 2021). The way in which the texts are structured and arranged based on their genre has implications for the readers' approach to comprehending the text and can subsequently compromise the learner (Graesser et al., 1991). The structure and the arrangement of written text and pictures in the passages can influence the way in which the learner reads and comprehends the text, hence it is important that the achievement booklets be layout verified so that all the learners taking part can get the same version of the achievement booklets but in their designated languages.

Comparable assessment translations and adaptations are crucial because they take place with the aim of generating equivalent and contextually adapted tests for all participating learners across different countries (Howie et al., 2017). The Organisation for Economic Co-operation and Development (OECD) in its Programme for International Student Assessment (PISA) defined adaptation as “an intentional deviation from the source version(s) made for cultural reasons or to conform to local usage” (OECD 2016, p. 3). The implication of the definition above is that the items or

phrases in the instruments are contextually modified on purpose with the intention of making them locally and culturally relevant to that country. This means modifying some of the items and phrases in the PIRLS achievement booklets so that they are familiar to the Grade 4 learners taking the assessments. For instance, the US uses the imperial measuring system while South Africa uses the metric measuring system. Learners will have better comprehension of measuring units such as centimetres. In the case of isiZulu, the term for centimetres is “Isentimitha” while the term for inch is “iyintshi”. However, this does not guarantee that the learners are able to make the connection that this is a unit of measurement, but they do deal with centimetres rather than inches in mathematics.

The purpose of adaptation is to find cultural and language equivalence between the SL and the TL. Adaptation is conducted to import the vital elements of one culture to the other, that process is executed to ensure that the assessments are appropriate for the Grade 4 participants of a specific culture and language. Adaptation is the altering of a text’s words and phrases to suit or fit the local context. Ebbs et al. (2023) state that the NRCs are strongly advised to keep the adaptations to a minimum, but are allowed to make the necessary adaptations that reduce unfamiliar contexts and vocabulary that may possibly hinder learners ability to read and comprehend the passages and items. Roux (2015) mentioned that the adaptations are acceptable if the learners are not familiar with the terminology used, that may impinge the learners’ ability to read and understand the texts.

The IEA has developed technical standards and guidelines for its ILSA to ensure quality in instrument production (Mullis & Martin, 2021). Since instrument adaptation has to do with the instrument intentionally deviating from the original version, to make the instrument culturally relevant and to adapt to the normally used language in that country, the IEA has released the necessary guidelines to guide the process of adaptations. For instance, in the SL, the text refers to *miles* to measure the distance; however, when adapted to the South African language it is referred to as *kilometres* to make it relevant to the context without compromising comparability. To maintain assessment standardisation of adaptations across the participating countries, the TIMSS and PIRLS International Study Center released a list of specific examples of

adaptations that are acceptable and those that are not (Ebbs et al., 2023). The process of adaptation is applied when there is a risk that the respondents would be either advantaged or disadvantaged if a straightforward translation were used. Subsequently, the translators are instructed not to use real names, such as cities or towns, to prevent students from being influenced by their knowledge and experiences of their national locations (Ebbs et al., 2023).

Adaptations made to the instrument are recorded on the national adaptation forms and sent to the NRCs for verification. Mullis and Martin (2021) reported that once the validity and reliability of the international version of the instruments are confirmed, procedures are put in place to establish the linguistic equivalence of national versions with an intention of gathering comparable data. The process of establishing equivalence takes place to ensure that the data are gathered from both the number of participants (all sampled participants) and their linguistic ethnic and cultural heterogeneity. For effective comparison to take place, all sampled participants including their difference in language and social aspects, are accounted for.

2.7 SAMPLING DESIGN

To provide valid and reliable measurement of reading trends of Grade 4 learners, PIRLS makes use of rigorous school and classroom sampling techniques. PIRLS 2021 made use of a sampling design called Stratified Two-Stage Cluster sample design (Wry & Mullis, 2023). The first stage of the design dealt with the sampling of schools and the sampling of classes in the school was categorised as the second stage of sampling. Stratification is concerned with the organisation of schools in the target population into groups that have similar characteristics such as school types and demographic region. Stratification variables include aspects such as language of instruction, school quintiles and socio-economic indicators. Each country's NRC is tasked with developing a national sampling plan for its national population and thereafter applying the PIRLS sampling techniques to obtain a sample that is nationally representative of the population tested. During the national sampling plan process, the NRCs are required to consult the PIRLS sampling experts. A series of forms are completed by each country's NRC providing information about the country's national target population, coverage and exclusions (Wry & Mullis, 2023). The proposed

sample data is sent to Statistics Canada and IEA Hamburg before field-testing begins. Precision of estimates of learner achievement are of utmost importance, consequently, there are standards and guidelines that each country should follow to achieve PIRLS sampling precision and sample size (Wry & Mullis, 2023).

In the first stage of sampling, schools are sampled nationally to be representative of the national population fitting the PIRLS requirements of the national sample. Schools are sampled with probabilities proportional to their size from the sample frame provided and containing the eligible learners. The sample frame is organised based on the important demographic variables. It should be noted that each sampled school is assigned two reserve schools in case the sampled school declines to participate (Wry & Mullis, 2023). Once the school has agreed and given consent to participate in the PIRLS assessment, there is a move to the second stage. In the second stage, classes are selected with equal probabilities within the school. Furthermore, intact classes of learners are sampled rather than individuals, this kind of sampling is believed not to disrupt the day-to-day school activities because it takes the whole class rather than individuals from different classes. The sampling of intact classes in the second stage is necessitated by the fact that PIRLS pays attention to learner curricula and instructional experiences arranged on a classroom basis. At this stage of sampling, it should be noted that if the sampled class declines to participate, it is not replaced (Wry & Mullis, 2023).

2.8 MAIN STUDY DATA COLLECTION

The IEA has standard procedures for data collection that are implemented by the country's NRC team (Mullis & Martin, 2021). Johansone and Flicop (2023) explain that standardised assessment materials and survey operation procedures are adapted based on the previous rounds to ensure that the current (PIRLS 2021) study data collection is of the highest quality (Johansone & Flicop, 2023). Methods and procedures from the previous rounds are adapted and used to carry out the next round to ensure that comparable data is gathered, that there is consistency across PIRLS trends and minimal changes take place based on time and circumstances. For instance, in PIRLS 2021, six new assessment blocks were developed replacing the one released at the end of the previous cycles (Johansone, 2023). The implication is

that some data collection procedures are similar to the ones used in the previous cycles.

In June 2020, field testing activities for PIRLS 2021 were carried out in preparation for the main data collection (Johansone, 2023). The feedback from field testing was used to enhance the operation procedures and instruments for the main data collection. Thereafter, in October 2020, the southern hemisphere was scheduled to start with data collection and the northern hemisphere was scheduled to start in March 2020. During this time, data collection activities were disrupted by the Corona Virus (COVID-19) pandemic which caused delays that resulted in the administration of the assessments taking longer than usual (Johansone, 2023). Following field testing and improvement of instruments and data collection procedures and activities, the IEA PIRLS proceeded to main data collection. In South Africa, data collection took place from August to October 2021.

Below is a brief outline of the major activities that IEA PIRLS engage in when conducting the main data collection. These stages are coordinated by the NRC of each participating country (Johansone, 2023):

- Contacting the sampled schools and sampling the classes to participate.
- Overseeing of translations and preparing assessment instruments. The second stage of the process includes the translation and layout verification procedure of the assessment instruments.
- Managing PIRLS assessment administration, followed by the scoring of the constructed response items.
- Creation of PIRLS data files (Johansone, 2023, p. 6).

PIRLS 2021 made use of two methods of data collection. Firstly, the questionnaires for background information to gather data that would contribute to the interpretation of the achievement results (*cf.* Howie et al., 2011; 2017). Secondly, the achievement booklets were used to test learners' reading abilities through the use of passages and items.

The current study is particularly interested in the achievement booklets used to gauge learners' reading abilities. In total, eighteen passages, nine informational and nine literary were used. The criteria used to select the PIRLS 2021 passages were decided upon during the first NRC meeting held in Hamburg in 2018. Criteria included aspects such as suitability of the content, passages be well written in terms of depth and complexity and avoiding bias in cultural differences. Thereafter, a discussion board was created to extensively review the chosen passages to determine their appropriateness, translatability and cultural suitability. The process of reviewing was carried out by the Reading Development Group (RDG) and NRC (Wry & Mullis, 2023). As previously indicated, PIRLS 2021 made use of a group adaptive assessment design, which encompasses an achievement booklet containing either difficult, medium or easy passages (Mullis & Martin, 2021).

In South Africa, the PIRLS data were gathered by the CEA at the University of Pretoria. The IEA Hamburg cooperates with the NRC of each country to arrange for data collection operations and to verify data accuracy and consistency within and across the countries (Mullis et al., 2017). Data collection verifications are done through survey tracking forms (Johansone, 2023). After the data are collected, scored, reliability scored, quality assured, verified and accurately documented, PIRLS moves to the next stage of the study.

2.9 DATA CAPTURING, PROCESSING AND ANALYSIS

Post-main data collection, IEA Hamburg, in collaboration with the TIMSS and PIRLS Study Center, Statistics Canada and the NRC of the participating countries, engages in an extensive process of ascertaining the integrity of the PIRLS collected data and preparation of the international database that contain valid, reliable and comparable data (Cockle, 2023). There are standardised procedures released by the IEA to each NRC of the participating countries on how to clean the data, identify variables, linkages and context data before analysis can begin. The international database is only created after the data have been captured and processed. The capturing process of data includes data cleaning, processing and feedback sent to the outsourced company (Howie et al., 2017).

The IEA releases the Data Management Software (DMS) with manuals and codebooks that specify the information on the IEA Data Management Expert (DME) fields in each of the released data files. The DME software is developed by IEA Hamburg and utilised for data entry and data verification (Martin et al., 2017) to ensure validation checks and identify any inconsistencies in the data (Cockle, 2023). The codebooks released by the IEA define structures of the data to be entered and it contains information about the names, lengths, labels and missing codes of variables and values for nominal or ordinal questions (Cockle, 2023).

The IEA and the CEA collaborated to create templates for data capturing of all instruments and forms (Howie et al., 2017). The country's NRC outsources a company to handle the capturing of data from all the instruments and the forms used. During this stage of data capturing, the data manager is responsible for training all the data capturers on how the DME program works, including how to enter the data and quality control the capturing of the data. The golden rule stipulated by the IEA is that the national staff data capturers must enter the data "as is" (Cockle, 2023, p.5) without any alterations, interpretations, imputations or cleaning.

To check for reliability in data entry, the NRCs are instructed to engage in a process called double-data entry, whereby the same data is punched into the system by two data capturers (Cockle, 2023). Only 5% of each survey instrument was required to be entered into the system by a different person. Possibly identified systematic misunderstandings or mishandlings of data entry require appropriate remedial measures. The NRCs are tasked with verifying and validating the data and undertaking corrections if necessary (Cockle, 2023). A uniformed data cleaning procedure including processing the national data, is further undertaken by the IEA. During that stage, IEA Hamburg continuously consults with the NRCs. The rationale for data cleaning and processing is to certify that the data adhere to international formats, the instruments that collected data can be linked across different survey files and the information collected is an accurate and consistent reflection of each country (Cockle, 2023).

Data analysis can only take place once the international database has been created, cleaned, verified and approved by the IEA. The PIRLS data can only be analysed

using the statistical software prescribed by the IEA and CEA researchers. Considering the complexities and weights of the numerical data, it can be analysed using IDB-Analyzer in conjunction with SPSS (Foy, 2018; Howie et al., 2017).

2.10 QUALITY ASSURANCE PROCEDURES

The TIMSS and PIRLS Study Center, in collaboration with the IEA Amsterdam, are responsible for developing international quality assurance programmes to ensure that the data collection activities are of the highest quality, and they are standardized across the participant countries (Johansone & Flicop, 2023). Furthermore, the quality assurance programme also has steps on how to verify that the standard procedures are followed. IEA Amsterdam appoints International Quality Control Monitors (IQCMs) to oversee the quality assurance programmes by visiting the schools in each of the participating countries to observe the administration of PIRLS 2021. The members of the IQCM were contracted on the basis that they have no relationship with any member of the NRC, and they must be either retired schoolteachers, school inspectors or ministry officials. It is important to mention that the data collection for PIRLS 2021 took place in the midst of the global pandemic, COVID-19. Therefore, Johansone and Flicop (2023) note that the issues that were found relating to the COVID-19 pandemic which might have coincided with data collection, were addressed on a county-by-country basis.

To ensure quality across all stages, each contracted member or company receives extensive training regarding their role in the PIRLS study (Johansone & Flicop, 2023). The IEA has various quality assurance procedures implemented in different stages of the survey. For instance, IEA Amsterdam collaborated with the TIMSS and PIRLS International Study Center to conduct an international quality assurance programme that included school visits with the purpose of monitoring and reporting on the administration of the assessments in the field (Mullis et al., 2017). As part of the quality assurance programme, after scoring, a quality assurer is trained to quality assure the achievement booklets post scoring (Roux, 2020). The IEA has developed technical standards and guidelines for ensuring quality in instrument production (Mullis & Martin, 2021) which takes place to determine the reading abilities of learners against national and international standards (Howie et al., 2017). Therefore, ensuring quality in

instrument production and implementation of the survey is mandatory to obtain comparable data.

2.11 CONCLUSION

This chapter discussed PIRLS 2021 in great detail as the source of this present study. PIRLS, an ILSA aimed at gauging Grade 4 learner reading literacy ability, providing feedback to the participating countries and assisting in improving their policies on reading literacy. PIRLS has different reading purposes and processes of comprehension that guide the development and design of the achievement booklets under the reading framework. The assessment design allows the participating countries to contextually adapt the instruments, for South Africa that meant that 70% of the passages in the instruments were less difficult and 30% were difficult passages. Initially, the booklets are designed in English, which means that South Africa had to translate the PIRLS instruments from US English to ten South African official languages. At regular intervals, South Africa had to check in with the TIMSS and PIRLS International Study Center for translation verifications. Before testing could begin, the South African NRC team contacted the sampled schools to begin with the arrangements and make certain that the schools were willing to participate.

PIRLS 2021, the 4th cycle conducted in South Africa, occurred from August to October 2021. The NRC compiled a team to ensure that each stage of the data collection, capturing, processing and analysis of the collected data aligned with IEA standards. In addition, IQCMs were appointed to ensure that all the PIRLS international standards were followed, and comparable data were collected. PIRLS 2021 in South Africa was a rigorous process that aimed at assessed the reading literacy levels of Grade 4 learners.

CHAPTER 3: A REVIEW OF THE LITERATURE AND THEORETICAL FRAMEWORK

3.1 INTRODUCTION

This chapter offers a review of global and national perspectives of literature in relation to the equivalence in assessments and it links the chosen theoretical framework with the study. Reading literacy rates in South Africa are low to a point where it has been made the nation's aim to promote reading literacy as a vital skill in a child's development both as a learner and as a citizen (DBE, 2023; Motshekga, 2017). Measurement invariance should be given attention to ascertain assessment equivalence and validity if the assessments are translated and adapted into different languages. Test bias can result from different factors, but in the case of this research, it is proposed that possible bias in the assessments may be due to the translation of the achievement booklets, hence this study sought to examine the validity of assessments, investigating equivalence, functional and metric equivalence to establish the degree of equivalence in the PIRLS 2021 items of one selected informational story.

This chapter begins with a description of reading literacy in the South African educational sphere with a focus on the development of reading literacy (Section 3.2). The following sections relate specifically to translation, discussing issues pertaining to translations that could possibly result in item bias (Section 3.3). Section 3.4 offers a discussion of measurement invariance in relation to this investigation. Section 3.5 focuses on the validity of assessments and Section 3.6 discusses assessment bias while Section 3.7 reviews the concept of equivalence by differentiating between the different types of translation equivalence. The theoretical framework of the study is introduced and discussed in the penultimate section before the chapter concludes.

3.2 READING LITERACY IN THE SOUTH AFRICAN EDUCATIONAL LANDSCAPE

"Before the democratic elections in 1994, South Africa's education system was divided along racial line" (Wolhuter, 2017: p.60). At the time, race and language were used as

a proxy to divide the education curricula. Post-1994, strides have been made to ensure equal access to education for all. The system introduced a number of policies and reforms resulting in the current Curriculum Assessment Policy Statement (CAPS) taught in government school and the Independent Examination Board (IEB) curricula taught in independent schools. This means that South Africa had two education systems, the one is poorly resourced and mostly comprised of black learners and the other is comprised of good resources with better-performing learners (Fleisch, 2008). South Africa is thus considered to be one of the countries that has the most unequal school system in the world (Mohamed, 2020). In his work, Spaull (2013) has made it clear that the presence of these two systems produces bimodal performance in the learner population depending on which education system the learner is in. Mbiza (2018) suggests that the government should try to adopt the curriculum and models used by well-performing schools to even the playing field.

Mohamed (2020) has suggested that the presence of two curricula have resulted in the underperformance of some South African learners. In 2011, the Southern and Eastern African Consortium for Monitoring Educational Quality (SEACMEQ) III study ranked South African Grade 6 learners' reading fourth out of 15 African countries for the richest 25% of learners and 14th out of 15 for the poorest 25% (Rule & Land, 2017). This finding illustrates that the wealthiest section of South African learners performs relatively better than the poorest 25%. However, in SEACMEQ IV, according to the DBE (2017), South Africa showed signs of improvement in their overall reading scores when compared with SEACMEQ III. In the Annual National Assessments (ANA) conducted in 2014, Grade 9 South African learners in both Home Language (HL) and First Additional Language (FAL) were having difficulties in responding to items that require the use of their own words (DBE, 2014). Learners were not able to interpret a sentence or provide an opinion when necessary. These findings suggest that a dire reading situation exists even in higher grades, aligning with the results from the PIRLS cycles which have revealed that the majority of Grade 4 learners in South Africa cannot read for meaning (DBE, 2023; Howie et al., 2011; 2017). If a Grade 4 learner is not able to *interpret and integrate ideas and information* in text, this problem may continue in higher grades which has a negative across all learning areas. These findings depict

that reading literacy in the South Africa education landscape needs attention and strategies for improvement.

Taking into account the results of ILSAs and other national assessments, development of education has become a top priority. The South African government spent 15% of the total budget on education in 2016 (Mbiza, 2018) and the World Bank and UNESCO Institute for Statistics (2021) revealed that in 2019 South Africa contributed 18.4% of its GPD towards improving the education sector in an effort to improve learner performance.

3.2.1 Language in Education Policy

As previously indicated, post-1994, a number of policies were introduced into the education system in an attempt to address the inadequacies of the previous system. One such policy was the Language in Education Policy (LiEP). This is a policy introduced in 1997 by the Department of Education (DoE), as an attempt to introduce the different official languages of South Africa (DoE, 1997) and address the issue of the exclusion of South African indigenous languages prior to1994. Thus, the aim of the LiEP was to redress past inequalities in terms of language and to further ensure that all the languages are integrated into the South African curriculum (DoE, 1997). The integration of South African official languages meant that the languages will be taught as a subject and used across the curriculum as a language of instruction. The function of the LiEP in conjunction with the South African Schools Acts (SASA, 1996b), is to grant the Schools' Governing Body (SGB) the privilege of electing the schools' HL, FAL and most importantly the language of instruction (DBE, 1997). The LiEP plays a significant role in the teaching and learning of language in South Africa because it promotes the indigenous languages that were oppressed prior to 1994 and endorses the issues of the historical and multilingual situation in South Africa (Nugraha, 2019). Although the LiEP advocates for bilingualism and the importance of home language education, the language of learning and teaching (LoLT) in most public schools in South Africa is English despite the variety of languages present and the fact that the majority of the learners do not have sufficient proficiency in English. Nwammuo and Salawu (2018) report that in as much as South African indigenous languages are

increasingly being used, the non-indigenous South African language, English, remains dominant and used across grades and schools.

In South Africa, languages taught at schools are offered as either HL or FAL. In some schools, another language is offered as a second language. The Department (DBE, 2011) differentiates between HL and FAL, in a school where for example isiZulu is offered as the HL, it assumes the role of a language spoken at home and therefore is used as the school's HL, whereas a FAL assumes the role of a language taught in addition to HL. FAL is taught for communication purposes and to equip learners with more than one language. FAL is the language that is introduced to learners when they start schooling in Grade 1 as an added language; learners transition to using the language as the LoLT at Grade 4 (DBE, 2011). However, in the majority of cases, the school's HL does not match the language many learners speak at home. That is, at home a learner speaks for example Setswana, and then at school the language chosen as the HL is isiXhosa and the FAL is English. Although LiEP is a practical policy because it promotes the use of all languages, its implementation is questionable because there are mismatches in language practice at schools. This means that there is a discrepancy between what was envisaged by the policy and what actually happens in the classroom.

Coetzee-van Rooy (2018) made the claim that LiEP is not realised in practice. According to Coetzee-van Rooy (2018), there is a lack of congruence between the LiEP and the natural language practices of learners. The implication is that although policy was developed to include all languages in South African, there is lack of support and use of those policies to ensure that African languages are used (Beukes, 2009). Coetzee-van Rooy (2018) states that there is a lack of specific language plans to effectively implement the policy and there is the inability to move past the previously dominating language used during the apartheid era. Due to the gap between policy and implementation, Nwammuo and Salawu (2018) suggest that strategies such as a curriculum that encourages the teaching of indigenous languages, research immersion programmes, vital material resources and updating and modernising of indigenous languages could be utilised to bridge the gap between policy and implementation. They suggest that the media could play a vital role in creating public

awareness to raise awareness of the LiEP and the use of indigenous languages. The role that language plays in the development of reading is vital to ensure that learners are equipped with the vital reading literacy skills as they read to learn.

3.2.2 Reading Literacy in the South African Curriculum

Reading is a process of obtaining and making meaning by critically interacting with written language (*cf.* Chapter 1). UNESCO (n.d.) asserted that *literacy* is an integral part of the right to education, it is also the basis of lifelong learning, and it steers sustainable development. *Literacy* is thus the "process of making use of reading, writing and oral language to extract, construct, integrate and critique meaning through interaction with multimodal texts" (Frankel et al., 2016: p.7).

From the two separate descriptions of reading and literacy, it is clear that *reading literacy* is the active involvement with the text to extract, create, incorporate and evaluate the meaning of the text read. Reading literacy is defined by Mullis and Martin as:

> The ability to understand and use those written language forms required by society and/or valued by the individual. Readers can construct meaning from texts in a variety of forms. They read to learn, to participate in communities of readers in school and everyday life, and for enjoyment (2021, p.6).

As such, learners need to develop reading literacy skills to fully comprehend the passages they read. Learners need to understand the text and thereafter, create their meaning out of what they are reading. Learners need to integrate what they have read into their ideas about the text and then evaluate the meaning to make sense of it. Reading is one of the most important abilities that learners acquire and develop as they progress through their early school years. It can be used for recreation and personal growth, and it equips young children with the ability to participate fully in their communities and the larger society (National Centre for Education Statistics (NCES), 2020).

Reading is associated with everything that people do, which includes social situations and practices (Kirk, 2015). The skill of reading is crucial because the learner makes use of it on multiple occasions, such as academic reading to gain information and socially to learn about relevant news and scan for quick facts. Reading is regarded as a receptive skill (Frankel et al., 2016) implying that for new incoming knowledge to be received, it is most likely to be read and/or listened to. With reading, the reader critically engages with the passage to receive and comprehend the information, or the meaning communicated in the text.

The South African curriculum (CAPS) envisages that in Foundation Phase learners should learn to read a story and practise extra reading on a regular basis as it plays an important role in learning to read (DBE, 2011). Learners at Foundation Phase must learn skills such as decoding and understanding sentences in order to learn to read a story. Furthermore, five main components should be incorporated when teaching reading: “phonemic awareness, word recognition, comprehension, vocabulary and fluency” (DBE, 2011, p.14-18), but each of these components needs to be taught explicitly and practised on a daily basis. To either read at an academic level or social setting, a child needs to possess such skills. Learning to read is arguably the most vital skill in a child’s life (Resnick & Hampton 2009; Shea & Ceprano, 2017).

In addition to the decoding skills that learners need to acquire and develop, Govender and Hugo (2020) state that the cognitive skills of reading and writing should be mastered at a young age. Therefore, being able to read is key in a learner’s development, and it is important that these reading literacy skills be sharpened and enhanced for current and future use (Western Cape Government [WCG], 2020). Reading literacy skills are used across the curriculum, which makes it the foundation for all learning, that is, *reading to learn*. For example, learners make use of their skills when reading notes and/or texts for the other subjects they are doing in that grade and in the future, they will utilise their reading literacy skills in the workplace. Learners who acquire intensive reading skills tend to perform better in school (Kirk, 2015).

The CAPS document (DBE, 2011) allocates a maximum of five hours for a language offered as FAL and a maximum of six hours for a language offered as HL. However, this period of six hours is divided among the language skills that learners need to

develop, leaving the skill of reading with five hours per two-week cycle, resulting in two hours and thirty minutes per week. The allocated time is limited considering the stages of reading involved. Costa and Zezlina (2013) refers to three stages of reading, namely pre-reading, during-reading and post-reading, all of which need time in order to be guided by the teacher. It is important that reading be designated enough time for the learners to sufficiently interact with the texts and to practise reading and understanding. Perhaps the school curricula may also be a contributing factor in the low literacy levels in South African because of the limited access to reading hours and materials. Background information is crucial to reading thus the school curricula should be adapted to ascertain a continuous development of background information and exposure to reading materials in young minds.

Rule and Land (2017, p.1) have argued that South Africa has "lost the plot" in reading education despite the fact that South Africa has undergone several curriculum transformations since 1994, which have resulted in the implementation of numerous national, regional and international assessment tools with the intention of improving education. These include Outcomes Based Education (OBE), the National Curriculum Statement (NCS) and the Curriculum and Assessment Policy Statement (CAPS). Over the years, assessment tools like Annual National Assessment (ANA) and Early Grade Reading Assessment (EGRA) have been used to assess learners' proficiencies. International large-scale studies such as SEACMEQ and PIRLS have tested certain aspects of the curriculum; however, South African learners have shown signs of poor performance (DBE; 2023; Howie et al., 2017; Spaul, 2012). This argument indicates that regardless of the efforts (changes in the curriculum and various assessments) made by the DBE, there is still poor performance in reading literacy. Rule and Land (2017) suggest that to move past the reading crisis, emphasis should be placed on reading for comprehension or meaning in the text. However, it is important to note that reading is a complex activity.

3.2.3 Complexities in Reading Literacy

Comprehension is the ultimate goal of literacy. However, there are many barriers that alter learners' understanding of what they are reading. The complexity of words and sentences, which includes syntax, terminology and spelling, can act as a barrier to the

learners' comprehension of what they are reading (Resnick & Hampton, 2009). The complexity of word features aspects such as multi-syllabic words, low-frequency words, and unusual and long sentences. Secondly, conceptual complexity involves the difficulties and simplicities of the information in the texts that affect the learners' comprehension, which includes the density of the ideas in the texts and the conceptual references made. The complexities embedded in comprehension passages make it difficult for the learner to comprehend, especially so for Grade 4 learners who are still acquiring the language itself.

Reading is a complex skill that integrates multiple strategies to make meaning out of what the learner is reading, furthermore, it requires learners to recognise words in the text and comprehend what they mean (Shea & Ceprano, 2017; Resnick & Hampton 2009). The schema theory, associated with reading comprehension, suggests that reading is the process of decoding meaning in the texts *via* recognising words, letters, sentences, phrases and cluster letters, to make meaning of the overall text (Zhao & Zhu, 2012). The words and phrases included in reading passages assist learners in forming meaning of the text they are reading. The recognition of the words in the passages aid in making meaning of the whole text and ultimately comprehending the ideas embedded in the texts.

Many factors have an effect on the learners' ability to comprehend a text. The complexity could be the readability of the text, the levels of meaning or purpose, the structure of the text, the conventionality and the clarity of the language as well as the knowledge demands of the text (Department of Education Louisiana, n.d.). Complexity of text was noted in the PIRLS achievement booklets, where passages contained abstract ideas, figures of speeches and complex words that learners might have difficulty in comprehending[7] even though these are included in the CAPS curriculum. The PIRLS passages were meticulously selected based on a set of criteria acceptable

[7] A thorough analysis of the chosen passage in the achievement booklet is provided in Chapter 5.

to the participating countries. The criteria included aspects such as number of words, culturally sensitive, age and grade appropriateness (Wry & Mullis, 2023).

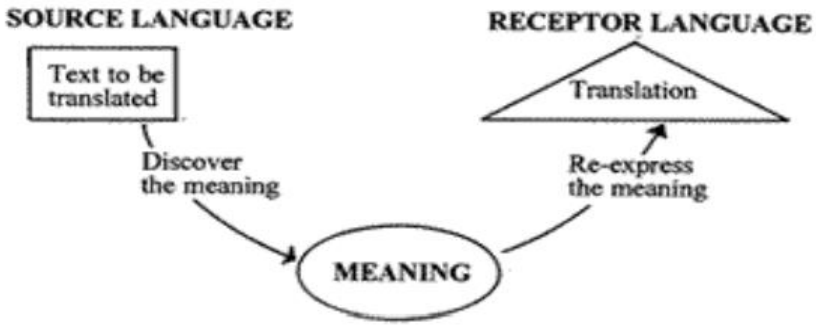
Within the complexity of text, vocabulary is another factor which influences understanding of text. The skill of decoding, which relates to the relationship between individual letters or letter combinations and words as an essential skill in reading fluency, comes into play. Decoding for learners could be problematic when the word used in the comprehension passage is not found in the language. For example, in PIRLS 2021, the term 'Hammerhead shark' was used in an information text. If a language does not possess all the words necessary for it to capture all its experiences, it 'borrows' or 'loans' words from other languages. This is particularly true with older languages and the development of new technology and discourses. The term 'Hammerhead shark' does not have a translation in African languages, which means that in translating from English to isiZulu, borrowed words had to be used, resulting in the term "*i-Hammerhead shark*" (hammerhead shark) in the achievement booklets. In the case of borrowing words, such as '*i-Hammerhead shark*', would mean that isiZulu learners might be challenged in decoding it and thus not understanding what it means or refers to. The learners' comprehension of the text could be hampered, and this possibly has an association with how they answer the items in the booklets. The process of decoding may be impeded by the words and phrases used, which could not be translated into the isiZulu language. To enhance learners understanding of the PIRLS passages pictures are included in the texts.

3.3 TRANSLATIONS

Translation is the science of moving a piece of text from Source Language (SL) to the Target Language (TL) whilst maintaining all the details incorporated in the SL (Tanrikulu, 2017). In simpler terms, translation is a process of converting a text from the language in which it was designed and developed, to another language; however, it should be seen as an effort of finding the equivalent meaning of the SL in the TL (Ronderos et al., 2021; Tanrikulu, 2017).

The proposed process starts with the text to be translated being evaluated and its meaning being discovered. The process proceeds to the main meaning re-expressed

through translation to the receptors' language. Figure 3.1 presents an illustration of the process of translation from the SL to the TL, proposed by Shiyab (2013).



(Source: Shiyab, 2013, p. 43)

Figure 3.1: An overview of the translation process

In translation, procedures are important for both the process of translation and the translator to objectively produce a target text similar to the original text both in form and content (Nida, 1969; Shiyab, 2013). Procedures are followed to systematically translate a text in the SL to the TL in a unique and specific way to ensure quality in the translation. Since the SL of the text has its own unique identifiers[8], they also need to be incorporated into the new translated version of the text (Shiyab, 2013). The meaning of the original text should not be lost during the translation to the target text, rather the meaning expressed in the TL must be equivalent to that of the SL (Chan & So 2017). The indication is that the reader of the source text and the target text must read the text in different languages and make the same meaning thereof.

It is challenging to maintain the structure in the target text because words do not have the same word length and tenses in the sentences and in addition, the grammatical issues differ by language, thus making it important for the translator to have substantial knowledge and fluency in both languages. Nugroho (2007) recognises the distinctiveness of the language by indicating that the system of rules in one language

[8] Language structures: idioms, syntax, semantics and morphology.

differs to the other language and the fact that some words and phrases cannot be directly translated into the TL from the SL. Although the meaning is at the centre of translation, it is vital to note that all cases cannot result in the meaning being the same due to factors such as the language structures, semantics, morphology, syntax and cultural differences between the two languages. In some cases, a word may express something that is understood in the SL but there is a similar word in the TL that equates to the expression. For instance, the SL is isiZulu, and the word is *"Ngilalele"* translates directly to "I am listening" or as the expression "I am all ears" in English as a TL. The above denotes grounds for different types of translations because the translator can choose which type of translation would be suitable in a particular instance (*cf.* Section 3.3.1). African languages have complex dialects, morphology and orthography, which makes the languages more complex when translation takes place (Mtsatse, 2017).

Languages are complex because they vary in their grammatical structures, word usage and difficulty and, as a result, the way in which one language in encoded and decoded differs (Berman et al., 2020; Solano-Flores et al., 2009). Translation depends on the decoding of meaning in the SL and re-coding of the same meaning in the TL and necessitates knowledge of how the meaning is shaped by cultural aspects in the SL and also how to phrase those in the TL (Solano-Flores, 2012; Solano-Flores et al., 2009). All the uniqueness and complexities of one language call for a need for adaptations of the assessments that are translated into different languages. The fact that a text requires a degree of adaptation entails that the two versions of the text cannot be identical. The features of the SL are not the same and cannot be the same as those of the TL. The variation in the text is controlled by the meaning of the test items, target population and TL. In other words, the text adaptations are based on the items, population and language to be selected.

3.3.1 Types of Translations

Among the types of translation, there is the direct translation procedure, a technique that encompasses *borrowing*, *calque*[9] and *literal translation*. Kembaren (2018) provides a clear distinction between borrowing and calque. *Borrowing* is considered to be an act of taking the word exactly as it is from the SL to the TL, on the other hand, *calque* is viewed as the translation of a phrase lexically and structurally. The *borrowing* method under direct translation does not translate the word to the TL, it simply takes the word as it is from the SL, which indicates that the borrowing technique takes the word in its natural form from the SL to the TL, with no translations or alterations made. During the translation process under *calque,* the focus is on getting the correct vocabulary and the form in which the word or phrase is written in the TL to preserve the unique language identifiers. The *literal translation process* is the process of translating word for word (Barbe, 1996). It is the translation of each and every word from the SL to the TL. Table 3.1 depicts examples of borrowing, calque and literal translation.

Table 3.1: Techniques used in direct translation

Technique	Source Language	Target Language
Borrowing	Hammerhead	i-Hammerhead
Calque	Weekend	Impelasonto
Literal	It is a school day today	Ilanga lesikole namhlanje

(Source: Different techniques utilised in the direct translation method. Compiled by the author)

In a direct translation study conducted by Hosseini (2021), the main findings revealed that direct translation is relatively overt[10], that transfers cultural elements from the SL

[9] A calque "(otherwise known as a loan translation): is a word or phrase taken from one language and translated literally, word-for-word, into another language" (Kembaren, 2018: p.98)

[10] Mismatch of explicit meanings of elements of the source and translation texts that include omissions, substitution or additions and the breach of TL systems like grammar and the acceptable level (Hosseini, 2021).

to the TL while the analysis of the findings revealed that these are no convert[11] errors in direct translations. This finding was reached through the use of Juliane House's 1976 translation quality assessment model. The aforementioned illustrates that certain translation techniques might work in certain languages, but it does not guarantee that it will work in other languages. Therefore, it is important to select a translation technique and theory that has limited or no errors when translated to the TL from the SL.

Indirect translation presents as another different kind of translation. Indirect translation is when word-for-word translation is not possible from the SL to the TL and when the SL cannot be translated to the TL without changing the grammatical structure and/or style (Karismawati, 2015). Indirect translation is also known as oblique translation which happens when the two languages are further apart and as a result, there is change to the text. Indirect translation may take place between English and isiZulu because the origins of these two languages and cultures are further apart. Rosa et al. (2017) states that the practice of indirect translation has been widely used in the history of translation.

Indirect translation is made up of different techniques, among which are *transposition*, *description* and *adaptation*. Karismawati (2015) differentiated between the techniques mentioned above. *Transposition* relates to the change made in the structure of TL without altering the meaning of the SL. Grammatical changes are made but the meaning is maintained as it is. *Description* is seen as when the term is not present in the TL therefore, the translator may resort into providing a description of that term in TL. *Adaptations* is concerned with the replacement of the cultural elements of the SL with the ones of the TL, which are similar (*cf.* Chapter 2).

[11] This kind of error results from dimensional mismatches that do not account for parameters such as field, mode and tenor during translation (Ehsani & Zohrabi, 2014).

Table 3.2: Techniques used in indirect translation

Technique	Source Language	Target Language
Transposition	Thank you	Ngiyabonga
Description	Hammerhead shark	Ushaka onekhanda elisasando
Adaption	Centimetres	Amasentimitha

(Source: Different techniques utilised in the indirect translation method. Compiled by the author)

3.3.2 Translation Theories

In the field of translation, many theories about translation have emerged to explain different approaches. Nida (1969) as a linguist, translation theorist and practitioner, proposed that anything that can be said in one language can be said in another language with reasonable accuracy by establishing equivalent points. Nida (1969) further recommended that translation should not entirely focus on stylistic details in a way that the procedure of translation will be difficult. Stylistic details include aspects such as the structure of the text and the positioning of the images. A language has its unique features that should be the focus during the translation procedure (Hoang, 2006; Nida, 1969). Features include techniques for linking clauses, unique patterns in the phrases, word building, and discourses embedded in the language to make it unique to others. Translators should be aware of these features and take them into account when engaging in translations. Lastly, in this approach proposed by Nida (1969), the meaning is more important than form, therefore the form may change to preserve meaning. Due to language differences, one-word expressions may be a phrase in another language. As an example, *"Sinethemba"* is a one-word phrase in isiZulu and in English, it is three words, "We have hope". In short, this approach advocates for meaning and not structure, as long as the meaning is maintained from the SL to the TL, then the structure is not an issue.

However, Nida's approach does not pay much attention to the fact that the TL has its own structural features that need to be acknowledged and respected during translation. In as much as the meaning is of utmost importance in this approach, aspects like sentence structure are also critical in the TL. Supposedly, that makes translation more complex because each language has its distinct components that

cannot be overlooked, for translators that make translations more difficult and challenging.

Van Staden (2006) in referring to the translation of PIRLS passages, pointed out that some of concepts and/or words are difficult to explain to translators. It seems that the translators understood the words in other contexts, but not in a way that they are expressed and used in African languages, which had an effect on the PIRLS translation stages and processes. This finding infers that a language has its own unique set of expressions and if they are to be expressed in another language, the meaning and context might not be the same, ultimately altering the cultural relevance of the expression. The above notion necessitates adaptations after the translations have been done. However, International Large-Scale Assessments (ILSA) such as PIRLS, have strict guidelines because of the need to maintain international comparability of the instruments.

Cultural differences also have a role in translation because some words and concepts in the SL or TL cannot be expressed in either the SL or the TL, be it concrete or abstract ideas (Nugroho, 2007). A concept can originate in one language due to cultural reasons and cannot be translated to other languages because it is abstract and native to that language (or simple because it does not have cultural relevance to other languages). It may be explained and rephrased into the TL, but the comprehension of that concept would not be equivalent to a cultural meaning when it is expressed in another language. Van Staden (2006) explained that most of the English expressions in the PIRLS passages were lost during translation into African languages because the expressions are not widely used in African languages.

Around the 1960s, Jiri Levy also proposed a translation theory that is different from other translation theories because it takes into account the translator, the translation process and the form of the translated text (Tanrikulu, 2017). The Levy approach acknowledges that the TL is also important - after translation, the literary work should not lose its literary value. Levy's approach recommends that the literary work communicated in the SL should also be present or transported to the TL. It is argued that in this approach, artistic aspects of a text can be determined by logic and those very same aspects should be translated and have the same value in the TL

(Jettmarova, 2012). They should have the same meaning and value as they are in the SL. However, it is difficult to achieve this kind of translation because languages are not the same and they do not have the same meanings and phrases.

Levy also viewed translation as a process that was made of up of three stages namely, *understanding*, *interpreting* and *reformulating* the text in the TL (Obdržálková, 2016). Jettmarova (2012) distinguished between the three processes involved in the Levy approach. In the first stage (*understanding*), the translator should not be too close to the work or text because contamination may occur. The translator should understand the process of translation and remain objective in the process of translation. Secondly, translation, as a process, centres around the recognition, *interpretation* and re-stylisation of the text. This stage of the Levy translation necessitates that the text be understood, paraphrased and format changed to suit the TL, but the meaning should not be distorted in any way. Lastly, it is *reformulation* that accounts for word principles of the text where characteristics such as phonetic patterns of the words, grammar and syntax are considered during the translation, which are regarded as the work of art. Translation should be consistent, meaning that the message found in the SL should be recreated in the TL (Obdržálková, 2016) and therefore, the texts become equivalent in both the meaning and form, and thus universal to all the text readers.

3.3.3 Translation Issues and Complexities in PIRLS Achievement Booklets

Alharbi (2017) claimed that when translating measures (constructs) from an original language to another language, the assumption with post-translation is that the measuring instrument assesses the same construct(s) in the SL and the TL. The claim suggests that the process of translation should not disturb or alter the constructs measured in any way or form. Furthermore, Alharbi (2017) sets a condition that if the assumption is not met, a translation error may have occurred and as a result, the measurements are no longer comparable across the different languages and cultures because they no longer assess the same construct. Thus, a possible translation error from one language to the other might invalidate the constructs measured in the assessments and then render assessments invalid and not comparable in the participating languages.

In brief, translations in assessments, especially ILSA, are critical and should be treated as such because a small error can result in the invalidity of the instrument used for measurements. Roux et al. (2022) emphasised that when an assessment is designed and implemented in different countries, it is crucial that the same construct is measured and that the achievement in the test only depends on the learners' proficiency in the subject or the topic area measured. Even so, the claims made above do not account for the differences in languages, language structures and language conventions. Again, there are different kinds of translation techniques used (*cf*. Section 3.3.1), which might contribute in terms of how the constructs are translated from one language to the other. As previously indicated, some words in the isiZulu PIRLS achievement booklets were borrowed from the English but with a prefix, i.e., *"i-hammerhead"* (*cf*. Van Staden, 2006). As the word still reads in English, it might be a challenge for learners who took the test in the isiZulu as they might not be familiar with the word.

Since PIRLS passages are multimodal, integrating written text and pictures, for the learners to make meaning of the passages with pictorial demonstration, an additional skill is required so that they can make the necessary connections. Here the learners' abilities in visual literacy are needed. There are numerous definitions of visual literacy, however, the University of North Carolina (UNC) (2020), provided an encompassing definition that looks at visual literacy as a set of standards that include interpreting the meaning and evaluating images. In the PIRLS passage *The Hammerhead Shark,* the learners might not have been able to decode 'hammer' and 'head' and then understand what a hammer is (a tool) and a head and equate that with the shape of the shark which is why it is called a hammerhead shark. It is crucial to note that words such as hammer might be foreign to isiZulu learners, and it becomes difficult for them to link such terms with their prior knowledge of the shapes of the sharks. Learners were required to equate the term Hammerhead shark with the illustration. The learners were required to use the contextual clues (in picture format) to develop understanding.

Peña (2007) explained that the translation of instruments and their procedures present challenges that can be seen as threats to validity alongside test-bias. Thus, ascertaining whether measurement invariance is a contributing factor in the case of

difference in the achievement scores is a prerequisite taking into account that the assessment instruments the learners are assessed with are translated from English into 10 South African indigenous languages (DBE, 2023).

3.4 MEASUREMENT INVARIANCE

The translation of the PIRLS 2021 instruments from English to isiZulu may have contributed to measurement invariance. As learners were tested in isiZulu, it might have meant that the translated passage did not have the same meaning for those tested in English, which would have meant that learners who completed the test in isiZulu would have experienced challenges in understanding the passage and answering the test items. The possible measurement invariance experienced in PIRLS 2021 may be a result of translation complexities such as the unavailability of words from the SL to the TL and the language differences. The current study was aimed at examining possible differential functioning in *The Amazing Octopus* passage and items, developed in English and translated to isiZulu, and to determine whether measurement invariance is a contributing factor in the difference in the Grade 4 learner achievement scores.

Measurement invariance has multifaceted aspects that can be investigated through the use of different statistics (Combrinck, 2020). Measurement invariance is concerned with the psychometric equivalence of a construct across different groups taking the same instrument (Putnick & Bornstein, 2016). Measurement invariance requires that the construct measured must be the same for all the respondents taking the assessment. If the construct has different meaning for the respondents across groups, then it results in measurement noninvariance (Putnick & Bornstein, 2016), where the construct cannot be meaningfully construed across groups. Combrinck (2020) states that if the instrument violates the condition of measurement invariance, the instrument items need to be investigated. If measurement invariance cannot be established, and respondents respond differently to the items, then they cannot be reasonable compared across the different respondents. If the means of the different groups are to be meaningfully compared, the instrument items should be stable, meaning that they should be “invariant” (Van de Schoot et al., 2015, p. 1).

The idea behind measurement invariance is everyone taking the instrument must arrive at the same meaning when reading the instrument. Van de Schoot et al. (2012) postulated that if measurement invariance can be established then the respondents would interpret the items the same way. Measurement invariance is a concern in cross-cultural research when groups tested speak different languages and the instruments are translated, as with PIRLS, a cross-cultural survey that uses translated instruments across different participating countries (Cheung & Rensvold, 2002). As measurement invariance can be attributed to a number of factors, Byrne and Watkins (2003) noted that item bias (measurement invariance) can be due to translation issues that might be present in the instrument.

This implication is that post-translation of the PIRLS passage, both English and isiZulu, learners must interpret the items the same and ultimately respond in the same way to the items. PIRLS 2021 results revealed differing English and isiZulu scores, which is a concern. It is for this reason that this study aimed to establish the extent to which items in the PIRLS 2021 examined passage are metric and functional equivalent. This study thus aimed to determine whether possible measurement invariance could have resulted in the difference in the Grade 4 English and isiZulu learners' reading achievement scores during PIRLS 2021.

3.5 VALIDITY IN PIRLS ASSESSMENT INSTRUMENTS

Valid and reliable measures are essential to the field that studies abilities, aptitudes and attitudes (Zanon et al., 2016). In relation to the current study, validity is crucial since this study, in a broader sense, is directed toward examining factors that might disturb validity in assessments. Kadir et al. (2019) are of the belief that if a test measures what it intends to measure, then it is a valid test. In other words, validity can be defined as the evaluation of whether the designed assessment is assessing what it initially intended to assess (Gareis & Grant, 2015; Siddiek, 2010). The term validity is used to ensure that the test accurately measures the intended knowledge, skills and competencies set when the assessment was designed. Validity is deemed as the most imperative quality of an assessment, consequently, it is important to ensure validity in an assessment (Gareis & Grant, 2015).

The importance of ensuring validity in an assessment is that when the test is administered to learners, the assessment processes produce data in the means of learners' responses to items that are used to make conclusions about their learning. The data is then interpreted so that it can infer information about the learners' level of knowledge and skills (Kubiszyn & Borich, 2013; Killen, 2010). In the case of PIRLS 2021, it assesses Grade 4 learner reading literacy. Thus, it is important to ensure validity in an assessment so that valid inferences can be made using the assessment data as evidence.

Scores obtained by the learners on a test should not be due to factors other than those the test intended to measure (Roux et al., 2022; Basterra et al., 2011). Reinforcing this idea, Roux et al. (2022) maintains that an ILSA taken by participants from different countries should measure the same construct and the performance of learners must rely on their proficiency in the subject matter tested, not any other means. It is, however, crucial not to shy away from the fact that learners' scores can be due to many factors that may include threats to validity (Fraenkel et al., 2019; Basterra et al., 2011).

Assessing is a matter of making use of data to establish evidence of learning, therefore a question can be asked about how it can be decided that the evidence is valid (Chen, 2010). Validity can be investigated through the different types of validities to warrant the degree to which the assessment has validity. The validity of an assessment is a function of multiple facets, where each facet contributes to the overall degree of validity (Gareis & Grant, 2015).

Specifically for this study, construct and content validity were explored to determine the degree of validity of the PIRLS 2021 instruments using *The Amazing Octopus* passage and items. Taherdoost (2016) posits that content validity is recommended whereas construct validity is mandatory to be established in instrument generation.

3.5.1 Content Validity

Content validity is defined as the extent to which the items of the assessment are fairly representative of all the domains the test seeks to measure (Markus & Smith, 2010).

The definition provided conveys that for an assessment to be rendered content valid, the contents of the test must account for and be representative of all the skills and competencies meant to be covered and assessed in the test. Content validity is seen as a conclusion of measured performance based on the subject matter that is being assessed (Kadir et al., 2019; Oluwatayo, 2012). In the context of this study, it refers to the degree to which the achievement booklet passages and items measure the predetermined processes of comprehension outlined by PIRLS 2021 (Mullis & Martin, 2021). The idea is that post-translation, English and isiZulu learners must arrive at the same meaning when reading the passage and answering the items.

Content validity is the most crucial concept for the measurement of skills and knowledge that is frequently employed in evaluation studies (Kubiszyn & Borich, 2013; Siddiek, 2010). Content validity involves content relevance that pertains to the elements of the measured construct, being measured by appropriate subject matter and content coverage that are associated with the items in the test being representative of the subject matter (Ing et al., 2015). In relation to this study, it has to do with English and isiZulu learners having the same item content after translations and adaptations. To ensure content validity, PIRLS has several measures in place (Von Davier et al., 2023).

To establish content validity, the assessment items must be inspected to see whether they correspond with what the assessment intends to cover (Kubiszyn & Borich, 2013). As articulated earlier in this chapter, validity is a function of multiple facets. Content validity alone cannot warrant the validity of the assessment, and with that in mind, this study also examined construct validity.

3.5.2 Construct Validity

To comprehend the basic definition of construct validity, it is necessary to first understand what a construct is and what validity is. A construct is defined by Brown (2000:9) as an “attribute, proficiency, ability and/or skill that happens in the human brain and is explained by theories”. As mentioned earlier in this chapter validity is when an assessment assesses what it intends to assess. Thus, construct validity can be seen as how well the constructs covered in the instrument are measured by varying

collections of associated items (Maree, 2019). Kadir et al. (2019) articulated that to claim that a test is construct valid, the test must indicate that it measures the ability it is initially designed to measure using a series of related items.

To illustrate the above, if reading proficiency is a construct, a test about reading proficiency is rendered construct valid if there are a set of items specifically designed to measure learners' reading proficiency. The PIRLS 2021 instrument aimed to test learner reading literacy abilities through different kinds of processes of comprehension (*cf.* Table 2.2). For instance, 20% of the total items included in the achievement booklets are specifically designated for the standard of *focusing on and retrieving explicitly stated information*, whereas 30% of the items are focused *on interpreting and integrating ideas and information*.

According to Killen (2010), when the assessment measures a specific construct, that assessment must give learners several attempts to demonstrate their understanding of the content. For a specific construct to be adequately measured, more than one item must be formulated to afford the learner multiple opportunities to demonstrate their abilities of the measured construct. The focus of construct validity is determining how many items need to be designed to assess a specific skill in order to clearly evaluate whether the learner possesses that ability or not. In the case of this study, it infers that for PIRLS to adequately measure, for example, the *Interpret and Integrate Ideas and Information* process of comprehension, for both English and isiZulu learners, all items formulated towards measuring that construct must function the same post-translation of the achievement booklets.

Post-translation, the assumption is that the same constructs are still measured the same in both English and isiZulu. If the items show signs of different functioning from analysis, then it may be concluded that the items are no longer measuring the same construct (Combrinck, 2020) or the measurement in the achievement booklets is not equivalent after the translation procedure. Therefore, the invariance can possibly explain the discrepancy in the learner achievement scores. This type of validity is necessary for the standardisation of the assessment (Maree, 2019). When an assessment is standardised, it means that all the test takers are tested using the same materials, under the same test conditions and the scoring protocols are the same

(Sireci, 2020). The standardisation of assessments is done to ensure validity and equivalence in the assessments for everyone taking the test. To check for construct validity, multiple sources of evidence must be produced to attest that the test measures the intended construct/s as well as evidence that it does not measure irrelevant constructs and the assessment is not biased.

3.6 ASSESSMENT BIAS

Assessment bias is measurement error where irrelevant constructs are measured (French, 2014). Test bias negatively influences test equivalence, consequently reducing the comparability of the test scores. It takes place when a construct measured is not the same across different groups (Aegisdóttir et al., 2008). The content of items differs in the languages of the test. For example, perhaps the PIRLS processes of comprehension are measured by different item content in English and isiZulu, inferring that the language post translation might not be equivalent, therefore, possibly prompting different learner understanding and item difficulty. Test bias prompts different styles of responding from the learners because the test is no longer the same for everyone taking it. The above discussion raises questions about the test being content and construct valid. Test bias can result from poor translation which in turn reduces the comparability of the scores (Aegisdóttir et al., 2008) and the validity of the assessment instruments. Subsequently, it is important to check for different types of translation equivalence post-translation of the assessment.

3.7 Translation Equivalence

In ILSAs, it is mandatory that the texts across different languages be equivalent to each other, both in difficulty and understanding of the learners. Kim et al. (2003) proposed that if an instrument used in a cross-cultural study is not equivalent, the validity of the results will be questioned. In ILSAs such as PIRLS, metric and functional equivalence must be established, which raises the issue of the importance of establishing equivalence in different versions of the assessment.

Affrman (2010) viewed equivalence in translations as a term that denotes the relationship between the original version of the instrument and the translated version

that allows the translated version of the instrument to be seen as the identical translation of the source text. Nonetheless, it must be noted that the above definition acknowledges that the two versions cannot be identical, but optimal and maximal resemblance is of utmost importance. According to Aegisdóttir et al. (2008), the term 'equivalence' focuses on the question of comparability of the test across different groups. In short, the term equivalence means that all the versions of the test instruments should be the same in almost all areas after translation for comparability to take place. For this study, it infers that both the English and isiZulu achievement booklets should be similar to each other after the translation and adaptation of the instruments.

Jones and Kay (1992) provided suggestions for reaching language equivalence stating that the method of translation must be matched with the goals of translation and test for psychometric properties. Aegisdóttir et al. (2008) emphasised that making use of a suitable translation methodology is important, as it increases equivalence in the translated instruments, thus limiting issues such as mismatches in the instruments.

Equivalence is closely associated with validity, thus making equivalence a necessary prerequisite for test validity (Affrman, 2010). Peña (2007) postulated that metric and functional equivalences are part of the aspects that need to be paid attention to when translations to different languages are concerned. Subsequently, the use of different types of equivalences such as metric and functional equivalence in this study could determine the equivalence (validity) of the PIRLS 2021 items and passage post-translation.

3.7.1 Metric Equivalence: Item Difficulty

Metric equivalence is concerned with the psychometric properties of the items used to measure the same construct in varying groups (Aegisdóttir et al., 2008) and is crucial in assessment instruments translated into different languages and in instrument adaptation from one language to the other (Peña, 2007). Peña (2007) regards metric equivalence as the difficulty of a specific item expressed in two or more distinctive languages.

Metric equivalence is defined as the quantitative method of assessing the cross-cultural equivalence of the translated instruments and is essential for examining construct validity (Kim et al., 2003). It can also be used to assess the validity of findings gathered from the translated instruments. Validity is assessed by examining the patterns of psychometric properties in the cross-cultural data from both versions of the instrument. Metric equivalence is determined by looking at the data gathered from the two instruments belonging to two different languages. Essentially, it is the equivalence of the item difficulty after the translation to the TL to determine the validity of the items. In this type of equivalence, it is crucial that an indicator indicates the same value after translation (Andersson & Osterling, 2014). The implication is that the measured process of comprehension must be the same for all test takers.

In ILSAs, it is mandatory that all the different languages texts be equivalent to each other both in difficulty and understanding of the learners. Kim et al. (2003) proposed that if an instrument that is used in a cross-cultural study is not equivalent, the validity of the results will be questioned. Affrman (2010) stated that it is ideal that the two forms of the test (assessments translated into different languages) be similar to each other. It is, however, acknowledged that “complete statistically identity” (Affrman, 2010, p. 39) is hard if not impossible to attain. Thus, items in the assessment might discriminate against learners’ different abilities but not to a point where it is discriminating against one group and favouring the other.

It is a common practise for ILSAs, specifically PIRLS, to develop instruments in English and then allow the participating countries to translate the instruments into their national languages so that they can participate in the assessment. Therefore, a question may arise about how valid the items are post-translation. To answer this question, this study examined the metric equivalence of the PIRLS 2021 *The Amazing Octopus* passage and its items to determine the equivalence of the items.

3.7.2 Functional Equivalence: Instrument Behaviour

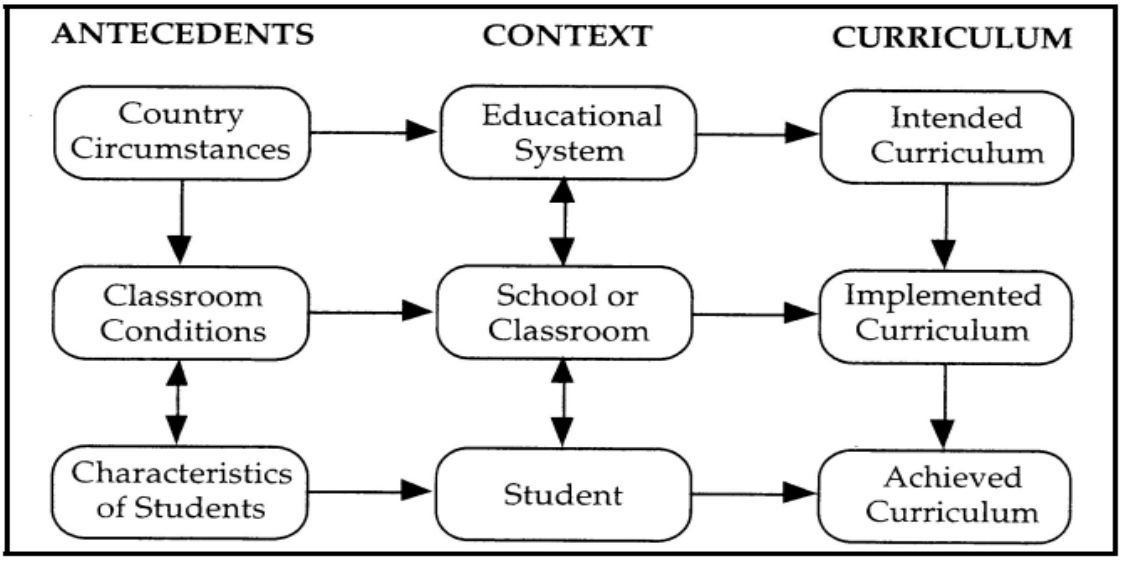
Functional equivalence has to do with the function and the behaviour of the items in different cultures or languages (Aegisdóttir et al., 2008). Peña (2007) defined functional equivalence in a way that implies that the translated instruments will elicit

the same behaviour in various languages to which the assessment is translated. Functional equivalence deals with the concept of item behaviour serving the same role or function in different languages (Kim et al., 2003). In the case of this study, it means that the same passage and items should behave the same way in English and isiZulu during testing. It was proposed by Aegisdóttir et al. (2008) that if a similar function (construct) has different behaviours in different groups then it cannot be used for comparability.

When translations take place, there is a possibility of incongruity in the meaning of the translated texts, therefore, content validity is threatened (Peña, 2007). Functional equivalence tests attend to such cases by examining whether the instrument item content still examines the same construct in different languages. Thus, the aim of functional equivalence is to ensure that the instruments are equivalent in item behaviour in the various languages to which the assessment instrument is translated and there are no signs of mismatches. If some kind of mismatch is found, it could be seen as a threat to validity of that particular test.

3.8 THEORETICAL FRAMEWORK

This study utilised the *Curriculum Process Framework* developed by the International Association for the Evaluation of Educational Achievement (IEA) (Mullis et al., 2007) This framework links the key role components in an educational system with regards to evaluating an education system (OECD, 2005) and exists at three levels, namely, the (a) *intended curriculum*: at this level, the curriculum is specified by those in charge of the education system, (b) the *implemented curriculum*: what actually happens in the classroom and the (c) *achieved curriculum*: this stage has to do with what was achieved from the learning experiences (OECD, 2005). This framework is aimed at linking what is envisaged by the educational system and what is implemented. Thereafter, it evaluates the learning experiences of learners through their performance in assessments. However, it is important to note that the levels present in the curriculum side of the framework do not exist in isolation, they are influenced by the antecedent and contextual factors operating within the systems, classrooms and learner level (OECD, 2005). Figure 3.2 shows the Curriculum Process Framework.



(Source: OECD, 2005, p. 9).

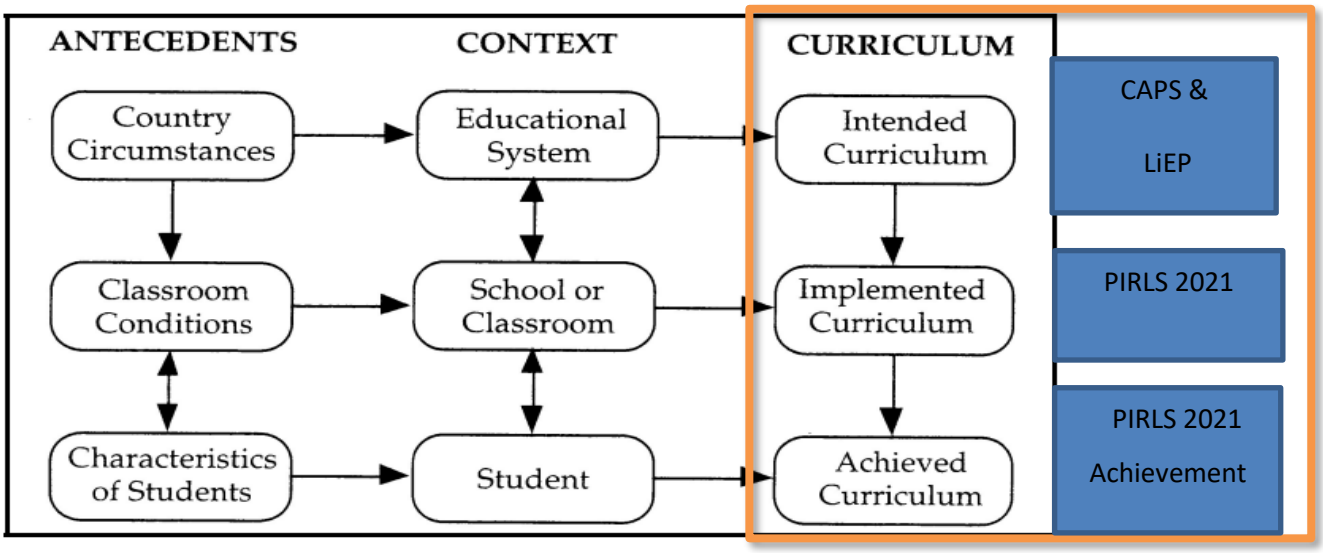
Figure 3.2: The curriculum process framework

Antecedents are considered to be the background conditions that might influence other variables (Kaya & Ok, 2020). These are background factors that are taken into account before the stages of the curriculum take place. The background factors influence the design, implementation and evaluation of the curriculum. The factors include the country's circumstances such as the needs of the country that the curriculum may address and resources and funds to be allocated towards that curriculum. Classroom conditions include the resources in the classroom and the teacher's readiness to implement the curriculum. Characteristics of the learners include their willingness to learn the curriculum, and their abilities, including their learning styles.

Context is defined by Null (2011) as an environment in which the curriculum is going to take place. This takes into consideration the education system within which the curriculum is going to be implemented, the school and classroom where the curriculum will be delivered and the student who receives the contents of the curriculum.

Although the framework consists of three levels (antecedents, context and curriculum), for the benefit of this study, only three sub-sections were utilised under the *curriculum*

level of the framework. The *intended curriculum* is paired with the aims and objectives envisaged by the South African Curriculum known as CAPS (DBE, 2011) and the LiEP (DoE, 1997) documents. The *implemented curriculum* is aligned with the PIRLS 2021 assessments whereas the *achieved curriculum* is linked with the learners' PIRLS 2021 achievement scores. Figure 3.3 shows the adapted section and sub-section of the Curriculum Process Framework that are applicable to this study.



(Source: OECD, (2006, p. 9).

Figure 3.3: Theoretical framework adapted from the Curriculum Process Framework

The *intended curriculum* is paired with the LiEP (DoE, 1997) which stipulates that learners should be taught in their HL – one of the 11 official South African languages. Schools were granted the privilege of choosing their own LoLT to expose learners to their HL as much as possible. The Curriculum and Assessment Policy Statement (2011) clearly stipulates that the policy is aimed at providing proficiency at the HL level, which refers to the development of basic interpersonal communication skills necessary in social situations and the cognitive academic skills required for learning across the curriculum. At Foundation Phase Level (Grades 1-3), the focus is on the teaching of the four language components specifically, reading, writing, speaking and listening. The CAPS document is aimed at equipping learners with the necessary skills like reading, to function in everyday life and their academics, whilst the LiEP states that the curriculum (aims envisaged by the CAPS document) can be taught in any

language. In PIRLS 2021, sampled South African Grade 4 learners took the assessment in the LoLT that school was using in the Foundation Phase of Grades 1 - 3.

The *implemented curriculum* represents the PIRLS 2021 standardised assessment which gauges what has been achieved by the CAPS (intended curriculum). The CAPS curriculum (DBE, 2011) requires learners to have developed reading proficiency at a Foundation Phase level in their HL and PIRLS 2021 tested the reading abilities of Grade 4 learners across the 11 official languages after they have exited the Foundation Phase level (Grades 1 - 3). PIRLS 2021 designs the ILSAs in US English and thereafter, disseminates them to the participating countries to translate into their national languages. That is in accordance with the LiEP because the assessments are then translated into the LoLT of the school. However, 30% of Grade 4 learners indicated that they completed the PIRLS 2021 achievement booklets in a language that is not authentically theirs (DBE, 2023).

The *achieved curriculum* is coupled with the PIRLS 2021 achievement scores of Grade 4 learners who took the test in English and isiZulu. Here the focus is on investigating the difference in the English and isiZulu learner achievement scores, and potential bias of the items that may possibly explain the difference in the scores between the two languages.

The stages of the theoretical framework are interdependent, implying that one stage is influenced by what happens in the other stages. The achieved curriculum depends on what happens in the intended and implemented curriculum. The implemented curriculum is implemented when the intended curriculum is disseminated. The intended curriculum is evaluated through the achieved curriculum. For this study, that means the PIRLS achievement scores of learners depend on the reading proficiency of learners envisaged by the education system and the way in which the assessment was designed to test reading abilities and administered in the field. In other words, the achievement scores of Grade 4 learners were investigated through the means of examining the PIRLS procedure of designing the assessment, which is the translation of the achievement booklets, hence this study argued that possible measurement

invariance (that my result from the quality of the translation of the achievement booklets) may help explain the difference in the learner achievement in Grade 4.

The reason for this study to utilise this theoretical framework and only the curriculum side of the framework is that this study investigates the difference in achievement scores of learners (*achieved curriculum*) from PIRLS 2021. For this study to do that, it needs to firstly understand the curriculum associated with reading (*intended curriculum*), what was tested and how it was tested *(implemented curriculum*), which ultimately leads to the *achieved curriculum,* which in this case, is the Grade 4 learners' achievement scores. Secondly, this framework accepts the investigation of learner attainment by acknowledging that there are antecedents and contextual elements that must be taken into consideration when the achievement of learners is concerned, although this is outside the scope of this current research, it is worth mentioning. The theoretical framework allows this study to look into learner achievement scores by firstly recognising the *intended* and *implemented curriculum*, which leads to the *achieved curriculum.*

3.9 CONCLUSION

In conclusion, this chapter reviewed and discussed literature regarding reading literacy in the educational landscape of South Africa. This chapter also focused on translation equivalences and validity in assessments and outlined the theoretical framework that provided a lens for this study. Reading literacy is a fundamental skill that is utilised across the curriculum and is important to develop an understanding of the 'reading crisis' by looking factors which could pre-empt it, such as the translation of the PIRLS booklets into the African languages, which meant investigating the equivalence of *The Amazing Octopus* across English and isiZulu languages. Possible measurement invariance may be associated with the difference in scores in the mentioned languages which thus be a possible threat to the validity of the PIRLS 2021 assessment. Validity is important in an assessment because the intention is to produce valid results and consequently, drawing valid inferences. It is therefore pivotal to establish translation equivalence to attempt to reduce assessment bias. This inquiry is structured to utilise

the curriculum process framework model that takes cognisance of the intended, implemented and achieved curriculum.

CHAPTER 4: RESEARCH DESIGN AND METHODOLOGY

4.1 INTRODUCTION

The research methodology used in PIRLS 2021 was discussed in Chapter 2. This chapter outlines the research methodology of the present study. In light of the problem statement discussed in Chapter 1, the main aim of this research was to determine whether possible measurement invariance could have resulted in the difference in the Grade 4 English and isiZulu learners' reading achievement scores during PIRLS 2021. To address the research problem and achieve the aim of the current study, secondary research as the design was used as it enabled the researcher to use existing data. Chapter 4 describes and justifies the design and the methods by which the PIRLS data were collected, analysed and interpreted.

The chapter begins with an overview of the research paradigm (Section 4.2) before moving into the methodological approach employed to collect and analyse data and the research design that guided the study (Section 4.3-4.4) and a description of the research questions (Section 4.5). The PIRLS 2021 research sample and the reduced sample for the current study is described in Section 4.6. Section 4.7 discusses the PIRLS 2021 instrument that this study examines which leads into a section (4.8) which offers a detailed description of the data analysis and interpretation methods utilised. The final two sections present the methodological norms and ethical considerations before the conclusion of this chapter.

4.2 RESEARCH PARADIGM

The term 'paradigm' was presented by Kuhn in 1962 in the seminal work, *"The Structure of Scientific Revolution".* Kuhn defined paradigm as a unified cluster of fundamental concepts, variables and problems tied with corresponding methodological approaches and tools (Shah & Al-Bargi, 2013). A paradigm is a set of assumptions that underpins a research project and includes world views and a framework that plays a critical role in guiding the research process (Sefotho, 2018). This means that the research paradigm and research methodology should be intertwined and linked to make a feasible study. Although there are a number of

paradigms used in research, the positivist paradigm was deemed appropriate to provide the epistemological lens to view this study.

The positivist paradigm was first introduced by a French philosopher, Auguste Comte, as the scientific method of investigation that includes experimentation, observation, logic and reason (Kivinja & Kuyini, 2017). This paradigm maintains that the data be gathered, analysed and interpreted objectively without the researcher's subjectivity, aspects which relate to quantitative methods of research. This notion is also articulated by Rahman (2016), that quantitative research is grounded on the positivist paradigm of measuring variables. The objectivity in this paradigm makes the findings valid and reliable to make scientific inferences and the findings can be generalised to other contexts with similar qualities as the study.

In this paradigm, the scientific methods of research are adopted as a means of data generation and data are gathered through the use of numbers that can be statistically analysed (Creswell & Plano Clark, 2007). This paradigm deals with quantities and quantifiable observed phenomena. It aligns with this study because this current research quantified the observed phenomena (potential bias and/or measurement invariance) using mean, item percentages, raw scores and Rasch differential functioning analysis outputs to determine whether possible measurement invariance could help explain the difference in learner achievement.

In the positivist paradigm, the researcher is required to use the deductive approach, an approach that works from general to specific (Cohen et al., 2011; Creswell & Plano Clark, 2007). The deductive approach is advocated in quantitative methods of data collection (Alomari, n.d). For this study, using the PIRLS 2021 data is beneficial because the current research can make generalisations about the population regarding the phenomenon being investigated.

The utilisation of this paradigm is justified by the fact that the data are quantitative and assisted in answering the sub-questions of this inquiry as indicated below.

1. How do the overall Grade 4 English and isiZulu learners' reading literacy achievement scores differ on PIRLS 2021?

2. To what extent can the difference in achievement be explained by possible bias (measurement invariance) between English and isiZulu responses during PIRLS 2021?
3. To what degree are the items functionally equivalent post-translation?

These sub-questions require numeric data, hence the appropriateness of the positivism paradigm. With the positivist paradigm grounded in quantitative methods and vice versa, it rationalises the use of positivism in this inquiry.

4.3 METHODOLOGICAL APPROACH

Methodology refers to a systematic theoretical analysis of the methods to be applied in research (Igwenagu, 2016). The methodological part of the research gives an overview of how the research problem is to be investigated. The positivist paradigm of this study requires quantitative methods of research; thus, this study makes use of a quantitative research approach. As such, quantitative methods of gathering and analysing data were employed to present one world view of the phenomenon under study.

Quantitative research, used in both natural and social sciences, has to do with quantity, entailing that an object, phenomenon or research variable can be described in quantity or in a way that can be counted (Mishra & Alok, 2017). Apuke (2017) defined quantitative research as the procedure of studying and explaining a phenomenon under investigation using numerical data systematically. The quantitative research method is viewed as a process of collecting and analysing numerical data, with the purpose of finding patterns, relationships, making predictions, maximising objectivity, replicating the inquiry and generalising the findings to the target population based on the selected sample (Cohen et al., 2011; Conrad & Serlin, 2011). Statistical methods are employed to make sense of what is investigated and thereafter refute or accept the hypothesis (Spicker, 2017).

The most important aspect of this approach is the researcher's relationship with society and the way in which results are analysed and interpreted (Gronmo, 2020). In this method, the relationship between the researcher and the researched phenomenon

is one and labelled as an “objective observer” (Creswell & Plano Clark, 2007, p. 39). The researcher must be distant from the research and therefore report truthful findings because they do not influence the events under study.

Quantitative research provides robust findings, which allow the extensive explaining and answering of the research questions (Pham, 2018). Quantitative methods take into account a secondary analysis of data (Cohen et al., 2011) and in this current study, the PIRLS 2021 numerical data assisted in addressing the research questions by providing a robust view of the phenomenon under study. This inquiry is a secondary analysis of the PIRLS 2021 data where data were analysed to investigate Differential Item Functioning (DIF) patterns in the items across English and isiZulu.

4.4 RESEARCH DESIGN

Research design is a basic plan that guides the research and ensures all the components of the research aligns. Punch (2011) described a research design as a fundamental plan for a research project that includes the strategy of research, theoretical framework, questions of what is to be investigated, the tools, together with the procedures of collecting, analysing and interpreting empirical data. The choice of the design to be used lies in the methodological approach selected to be followed in the study. This study made use of a secondary analysis research design, which according to McMillan and Schumacher (2014) is a method of re-analysing the already collected data for the purpose of answering the new research question(s) that might have stemmed from the original data. Johnston (2014) further clarified that secondary research design is an inquiry into what is already studied with the purpose of learning what remains to be learnt about the preliminary area of investigation. A secondary analysis research design is associated with the fact that it is aimed at enhancing primary research (McMillan & Schumacher 2014).

PIRLS 2021 (as the primary investigator) investigated the reading abilities of Grade 4 learners using instruments, specifically achievement booklets, and found that isiZulu learners achieved 267 (SE=6.5) score points whereas English learners obtained 382 (SE=14.5) score points, placing isiZulu learners approximately 2 years behind in schooling years. The approximation is made based on the difference of the scores.

This means that isiZulu Grade 4 learners are underperforming when compared to the learners who took the test in English. Based on the PIRLS 2021 results, the current research aimed at looking at possible measurement invariance (that could be due to translation of the achievement booklets) and the difference in Grade 4 learner achievement, with the intention of investigating learner performance by identifying possible factors that may explain the difference in the achievement scores.

4.5 SUMMARY OF RESEARCH QUESTIONS

This study used the PIRLS 2021 South African database in an attempt to address the main research question posed by this study: *To what extent are The Amazing Octopus items equivalent across English and isiZulu language groups post translation?*

In order to address the main research question, three sub-questions were asked.

1. *How do the overall Grade 4 English and isiZulu learners reading literacy achievement scores differ on PIRLS 2021?*

Since the previous round of PIRLS showed discrepancy in the learner achievement (Howie et al., 2017), sub-question one was formulated to investigate learner performance in PIRLS 2021 and how it differs in terms of the mean scores of the mentioned language groups. Thus, descriptive statistics provided by Statistical Package for Social Sciences (SPSS) in conjunction with IDB-Analyzer developed and recommended by the IEA to analyse PIRLS data, were used. The utilisation of the software programs mentioned was compulsory because they have the capacity to handle and analyse the ILSA PIRLS data sets (Foy, 2018). The descriptive statistics for this sub-question focused on the raw scores, mean scores, the percentages of those who were able to get full marks, 50% and those achieved less than 50%.

2. *To what extent can the difference in the achievement be explained by possible bias (measurement of invariance) between English and isiZulu responses during PIRLS 2021?*

The results of sub-question 1 led to the formulation of sub-question 2, that was structured to investigate DIF in the items to provide an explanation of the difference in

the learner achievement found in sub-question 1. Sub-question 2 looked at the possible item bias post translation that might have contributed to the difference in the achievement scores of learners who took the test in English and those who took the test in isiZulu. The purpose of this question was to determine whether the items reveal signs of discrimination against any language and item after the translations and that the same construct is measured in the two languages, thus determining construct validity. In order to address this sub-question, Rasch Measurement Theory (RMT) was used through the statistical program Rasch Unidimensional Measurement Model (RUMM2030). Different kinds of tests were conducted and the RUMM2030 program provided this study with outputs to determine the possible DIF of the items (Combrinck, 2020) administered in English and isiZulu.

3. *To what degree are the items functionally equivalent post-translation?*

Sub-question 3 was posed to investigate the item behaviour in the two language groups. Using RUMM2030 outputs, this study was able to determine the item function in the two languages through the use of test of fit details that provided individual item fit (Combrinck, 2020). Functional equivalence in this study dealt with the item content serving the same meaning for both English and isiZulu learners after the translation. This sub-question was necessitated by the fact that post translation, the items might not have the same meaning for the learners in the two languages, reducing content validity and prompting different responses that might have a contribution in the difference in the achievement scores.

Statistics from the sub-questions were used to ascertain the equivalence of the PIRLS 2021 items in *The Amazing Octopus* passage. Raw scores were calculated, means scores were compared via Analysis of Variance statistics (ANOVA Statistics). Item performance was assessed through the use of Rasch Analysis Item Characteristic Curve (ICC) graphs that showed whether the items were difficult or not difficult for a particular group, thus implying that item difficulty was equally or not equally dispersed and negatively/positively influenced the validity of the test (De Bruin, 2010) and reliability of the scores.

4.6 RESEARCH SAMPLE

The research methodology and the epistemological paradigm of this study support the use of random sampling procedures, and it is ideal that the sample be representative of the whole population investigated for generalisation to take place (Okeke & Van Wyk, 2015). With reference to PIRLS and this study, each NRC of the participating countries is tasked with developing a national sampling plan (frame) for its national population and thereafter, must apply the PIRLS sampling techniques to extract a sample that is nationally representative of the population under investigation (Wry & Mullis, 2023).

PIRLS 2021 made use of a sampling design called Stratified Two-Stage Cluster Sampling technique (*cf.* Chapter 2). In PIRLS 2021, Grade 4 learners were assessed across 11 official South African languages and across all nine provinces. The PIRLS 2021 sample, which was nationally representative, comprised 12 426 Grade 4 learners, stratified by language and province (DBE, 2023). The implication is that the South African PIRLS 2021 sample drew its participants from the 11 official languages and nine South African provinces. The sampled learners came from 321 schools, and the participation rate in Grade 4 was 97% after the replacement of schools that did not participate in the assessments.

As the current study focused on only two language groups, it only utilised two groups (English and isiZulu) within the South African PIRLS 2021 national sample. Subsequently, sample reduction took place to reduce the original 12 426 Grade 4 learners sampled to take part in PIRLS 2021. For the purpose of this study, only 505 Grade 4 learners were sampled due to the criteria of completing *The Amazing Octopus* in either English or isiZulu. Out of the 505 that completed *The Amazing Octopus* passage, only 411 scores were valid for Rasch analysis and the remaining 94 were extreme scores that were removed. Only 196 scores were valid for English language and 30 were extreme and only 215 for isiZulu language were valid scores and 64 were extreme.

4.7 PIRLS 2021 ASSESSMENT INSTRUMENTS

Chapter 2 provided a detailed description of the PIRLS 2021 data collection processes. Since this study takes the form of a secondary analysis, it made use of the PIRLS 2021 data. During the PIRLS 2021 data collection, achievement booklets and contextual questionnaires instruments were used to collect learners' performance in reading and gauge the learners' reading attitudes. However, this study only focused on the assessment instruments, also known as the test booklets. Specifically, this study examined one of the passages contained in the test booklets, namely *The Amazing Octopus*. Table 4.1 outlines how the items in *The Amazing Octopus* passage are distributed in the different processes of comprehension and the item type.

Table 4.1: Item distribution according to the item type, processes of comprehension and maximum score for The Amazing Octopus

Reading Purpose	Item No	Item Type	Process of Comprehension	Maximum Score
Acquire and use information	1	Constructed response	Make straightforward inferences	2
	2	Constructed response	Focus on and retrieve explicitly stated information	1
	3	Multiple choice	Evaluate and critique content and textual elements	1
	4	Multiple choice	Focus on and retrieve explicitly stated information	1
	5	Constructed response	Focus on and retrieve explicitly stated information	1
	6	Constructed response	Interpret and integrate ideas and information	2
	7	Constructed response	Make straightforward inferences	1
	8	Multiple choice	Focus on and retrieve explicitly stated information	1
	9	Constructed response	Evaluate and critique content and textual elements	1
	10	Constructed response	Focus on and retrieve explicitly stated information	1
	11	Constructed response	Focus on and retrieve explicitly stated information	1
	12	Constructed response	Make straightforward inferences	2

Reading Purpose	Item No	Item Type	Process of Comprehension	Maximum Score
	13	Multiple choice	Make straightforward inferences	1
	14	Constructed response	Interpret and integrate ideas and information	3
	15	Constructed response	Evaluate and critique content and textual elements	1
			Total points	20

(Source: PIRLS 2021 User Guide for International Database by Fishbein et al., (2023), Boston College, TIMSS & PIRLS International Study Center. Copyright 2023 by Boston College, TIMSS & PIRLS International Study Center)

4.8 DATA ANALYSIS AND INTERPRETATION

The quantitative method involves a systematic analysis of the phenomenon under study *via* mathematical or computational techniques in a numerical form that includes but not limited to statistics and percentages (Mishra & Alok, 2017). Quantitative research equates to the use of statistical methods of data analysis to establish the connection between what is known and what remains to be learned through that specific research (Creswell & Plano Clark, 2018). It is known through the PIRLS trend cycles that there is a 'reading crisis' in South Africa (DBE, 2023; Howie et al., 2006; 2011; 2016), and it is continuously being investigated by multiple researchers (including the present research) as to the possible factors or reasons for the poor literacy results (Roux, 2020; McLeod Palane, 2017; Mtsatse, 2017; Labuschagne, 2015).

This research specifically studies the possible measurement invariance as a factor that might explain the difference in the PIRLS 2021 achievement scores of learners who took *The Amazing Octopus* in English and isiZulu by making use of statistical methods of analysis that have the power to determine whether there is item bias in the assessment. In the quantitative analysis approach, the analysis of data is constituted by statistically analysing scores collected using instruments with the intention of addressing the research questions. For the current study, the analysis of the Grade 4 learners' reading literacy achievement scores was conducted to answer the research questions of this inquiry, as stated in Section 4.5.

The data used in this study had been collected and documented by the initial researchers. The data were cleaned, verified and ready for use. The process of data preparation was done through several stages, as discussed in Chapter 2. PIRLS 2021 assigned unique identifiers to the data; for example, each country in the PIRLS data has its unique code (Howie et al., 2017). The IEA distributes Data Management Software (DMS) with manuals and codebooks that specify the information on the IEA Data Management Expert (DME) to each participating country (*cf.* Martin et al., 2017; Chapter 2).

4.8.1 Overview of Statistics Used in this Study

This study is a secondary analysis of PIRLS 2021 data sets that could only be analysed using the statistical software prescribed by the original researchers. Therefore, this study made use of the IDB-Analyzer in conjunction with SPSS to analyse the quantitative data. Descriptive and inferential statistics were used to measure the deviation of the test scores of isiZulu learners from the test scores of English learners. *Descriptive statistics* were utilised to understand the basic layout of the data about the population and to make predictions about the parameters of the population or sample. *Inferential statistics* lie in descriptive statistics and their function is to generalise the findings to the population from the sample studied (Creswell & Plano Clark, 2007). This study began with the descriptive part of the analysis followed by inferential statistics before conducting a DIF analysis and equivalence testing.

Descriptive statistics were calculated using SPSS plug-in with IDB-Analyser to analyse the means and percentages of the data to provide an overview of the differences in the scores of English and isiZulu learners. *Descriptive statistics* were also calculated with the objective of understanding the PIRLS data investigating the number of learners able to obtain full marks, 50% and/or less than 50% for an item. The analysis indicated the percentages of the number of learners who were able to correctly/not correctly answer each item of the passage. Raw scores and overall mean scores per language were calculated by recoding the data, which were done manually, to provide a more renounced visualisation of the data sets. The descriptive statistics generated in this investigation were of great assistance in addressing sub-question 1.

Thereafter, *inferential statistical* analysis was performed with the aim of providing a more complete and deeper layout of the data and for the purpose of hypothesis testing to either approve or reject the null hypothesis.

- The null hypothesis of this investigation states that the mean scores of the learners who took the test in English and isiZulu are not different and therefore measurement invariance is not a contributing factor (H_0 = μEnglish = μisiZulu).
- If the null hypothesis is rejected, the alternative hypothesis states that the means scores of the learners who took the PIRLS 2021 assessments are different and measurement invariance might be the contributing factor (Ha = μEnglish ≠ μisiZulu).

Rasch analysis was conducted for inferential statistics and additional software, RUMM2030, was used to analyse possible item bias through the Rasch Measurement Theory (RMT), using Differential Item Function (DIF) techniques to measure metric equivalence/measurement invariance explored by sub-question 2. In simpler terms, this study made use of the Rasch Analysis model to examine the possible differential item function across the languages in order to conclude whether there is any bias in the items in *The Amazing Octopus* passage, either favouring or discriminating against the languages. Cronbach's Alpha was used to do summary statistics to assess whether the data fitted the Rasch model. Thereafter, reliability analysis for each language group was investigated. Subsequently, individual item fit statistics were generated to partly assist this study in accepting or rejecting the null hypothesis. Statistical outputs produced by RUMM2030 also assisted in addressing sub-research question 3.

And again, using RUMM2030 after it was seen which data fits the model, DIF was investigated. Using RUMM2030, the current research was able to discover and ascertain which items were functioning differently in the investigated languages groups. The statistical outputs produced by RUMM2030 provided an analysis summary of the data that includes ANOVA that displays a person factor ANOVA in each of the set of items. Outputs from RUMM2030 are represented graphically (*cf.* Chapter 5) using the ICC graphs.

DIF statistical techniques are methods that test for differences in the items administered to respondents of the same sample with little dissimilarities. These methods are informed by the principle that if different groups of examinees have roughly the same level of knowledge and skills, then they should perform relatively the same in the test items regardless of the differences in the groups (Andrich & Hagquist, 2015; Zumbo, 1999). In the case of this study, when applying the DIF notion, it means both the English and isiZulu groups have similar reading knowledge and skills (based on the fact that they fit the target population of the PIRLS trend cycles; that is, they are all Grade 4 learners). Therefore, they must perform similarly in the PIRLS achievement tests, their achievement scores should not reveal a large discrepancy. However, the PIRLS 2021 results showed a difference between the two languages, hence the investigation of DIF in the items of *The Amazing Octopus* passage and items. DIF is the investigation of bias in the test items with the aim of identifying whether the test items show any signs of item difficulty or item discrimination towards a particular group (Zumbo, 1999).

DIF analysis is a mandatory element in evaluating educational test validity, because through individual item analysis, DIF shows where the items function differently for the test takers and where equivalence might have not been achieved, resulting in item bias and inadequate validity in the test (Alavi & Bordbar, 2017). Item fit statistics were conducted to determine the validity of the item parameters and to detect DIF. The results from items showing bias can harm the scores. This brings the concept of validity into question, the test assessing the similar construct or the test being the same and assessing what it intends to assess in the given sample of the population.

Diaz et al. (2021) refer to another important trait of looking into DIF in assessments. They suggest that DIF addresses the validity of a test score's function and/or purpose because without validity, a test score is not useful. When the items in the test show signs of DIF, it becomes a threat to construct validity because the items reveal that they are no longer testing the same construct in the different groups sampled and tested. Bias in the assessments can result in systematic errors that distort the inferences made using the test scores (Zumbo, 1999). The implications are that the

assessment designers, researchers and administrators are retrieving incorrect facts and making misinformed judgements, policies and educational changes.

Killen (2010) discussed several aspects in assessments that should be paid attention, specifically in cross cultural assessments such as PIRLS:

- *Item discrimination:* items should discriminate but not to the point where they function differently for the different group of learners;
- *The number of test items:* an assessment should have adequate number of items that will sufficiently allow the learner to demonstrate a certain construct;
- *Item difficulty:* items in a test should vary in terms of their difficulty, ranging from easy, medium to difficult; and
- *The layout of the assessment:* it is proposed that the test should start with easy items and proceed to difficult as this will prompt learners to continue to answer to the questions rather than coming across the difficult items first (Killen 2010, p.353-354).

In nature, the Rasch Analysis DIF methods place the examinees of different groups according to their test scores and thereafter investigate whether the test items create problems for a specific group (Zumbo, 1999). For DIF results to be effective, it is crucial to take into account the magnitude of DIF effects. In the present investigation, through RUMM2030, the ANOVA *F*-values produced a rank order of the items according to the magnitude of their DIF. The RMT and its abilities to estimate were predominant in standardised assessments where the main aim was to be able to precisely infer what all the individuals who took the test know and can do at each determined level of ability. Rasch Analysis estimates are sample-free and test-free under varying conditions (Stemler & Naples, 2021).

The RMT model aligns the person and the items in a line, items are arranged based on their difficulty (Combrinck, 2018; Hagell, 2014). This model acknowledges that the items are different as are learners' abilities, hence it aligns both the persons and the items in a linear line and takes into account the difficulty of the items. Linacre (1994) stated that the belief about this model is that it is unidimensional because it has the power to measure one construct trait, inferring that scale items must represent a

common latent variable. The *traits* in question are complex to be observed accurately, and as a result, they must be inferred based on individual performance, implying that this model can analyse and interpret the data scores of individual learners. The reason for using this model is that it has the capacity to analyse the learners' scores taking into consideration different underlying factors.

4.9 METHODOLOGICAL NORMS

This study took the form of a non-experimental study, as a result, validity and reliability are of utmost importance. PIRLS has several measures in place to ensure the validity of its trends and its reliability because it has been taking place for several years internationally (Korsnakova et al., 2020). To ensure validity, PIRLS employs a series of quality assurance procedures to assure each step of the study from instrument production, pilot testing, main data collection to data analysis (*cf.* Chapter 2). PIRLS has strict guidelines and verification points during each stage of the study, each country National Research Coordinator (NRC) consults with the IEA at a specific interval of the study before moving forward. In South Africa, the Centre for Evaluation and Assessment (CEA) oversees most of the proceedings of the PIRLS assessment to make certain of the surveys' validity and reliability measures. The researcher of the current study can confirm the above since he was contracted to score, reliability score and quality assure the PIRLS 2021 achievement booklets. Before the scoring took place, he was extensively trained (using scoring manuals provided by IEA) on what to do at each stage of scoring and quality assuring the instrument post scoring. The implication is that each step of the PIRLS survey has a checkpoint to ascertain validity and reliability envisaged by the IEA and to ensure that the procedural manuals provided by the IEA are followed to the letter.

PIRLS also had several checkpoints and reviews by experts specifically outsourced during the design and development of instruments. The translation of these instruments was overseen by different experts to ascertain that they were not duly influenced, and they met the international standards set by the IEA (*cf.* Chapter 2). Furthermore, PIRLS made use of International Quality Control Monitors (IQCM) appointed to ensure that the PIRLS methods and procedures sent to the participating countries are correctly followed (Johansone & Flicop, 2023). The main responsibility

of the appointed IQCMs was to oversee the implementation of the PIRLS quality assurance programme.

This study adopted validity and reliability procedures used by PIRLS 2021 during the primary collection of data. This study interpreted and reported the data objectively and data were utilised only for the purpose of this study. To ensure validity and reliability in the analysis of this study, before the data were analysed, the researcher recoded the PIRLS data to fit the software used. Furthermore, the researcher ran the statistics twice, to ensure that the data files selected in the PIRLS data addressed the research question and to ensure consistency in the results of the initial analysis. It is crucial to note that the data presented in this study were collected during the midst of a global pandemic, COVID-19, and learners were not at school for a period of time (DBE, 2023). Based on the fact that this current study was based on PIRLS, only the PIRLS 2021 datasets were included, and any other datasets were excluded. To fully address the main research question and sub-research questions of the present study, only PIRLS 2021 datasets were deemed suitable.

4.10 ETHICAL CONSIDERATIONS

The researcher successfully defended the topic of this research in front of a panel consisting of two critical reviewers, one defence chairperson and two supervisors. The primary researcher of this inquiry obtained ethical clearance to proceed with the investigation and make use of the PIRLS 2021 data. Ethical clearance was obtained from the Ethics Committee at the Faculty of Education, University of Pretoria (EDU161/22) (*cf.* Ethical Clearance Certificate). The researcher sent a consent letter to the IEA to get approval and permission to use the PIRLS 2021 data sets for the purpose of this study (*cf.* Appendix A), and permission was granted. As the researcher also looked into the isiZulu translation of *The Amazing Octopus*, permission was also sought from the DBE (*cf.* Appendix B). The researcher had no direct contact with the Grade 4 participants, however, it must be noted that he was part of the PIRLS 2021 scoring team assigned to score and quality assure the PIRLS 2021 isiZulu achievement booklets. From the large PIRLS sample only the English and isiZulu languages group information was extracted to serve the purpose of this study. The data's integrity was kept at all times.

The raw data will be stored for 15 years, as required by the University of Pretoria. The data will be protected to keep the identity of the PIRLS Grade 4 learners' achievement scores unknown to the public domain, therefore unique identities will not be used during the discussion and interpretation of findings, as required by the Protections of Personal Information Act (POPI Act). The findings of this study will be presented or published by the University of Pretoria. The PIRLS original researchers will have access to the publication as they are currently based at the University of Pretoria, CEA. This study was condensed into an article submitted for review in a prospective scientific journal of interest and presented in a conference proceeding without compromising the integrity and the sensitivity of the data.

4.11 CONCLUSION

This study was underpinned by a positivist research paradigm that advocates for the use of objectivity and scientific methods of research. As a result, this study employed the quantitative methods of research guided by a secondary analysis research design. Data from the PIRLS 2021 main study were used in the secondary analysis. This investigation adopted the sampling procedures used by PIRLS 2021, which includes the sample size and the sample frame but followed a reduced sample. Data were analysed quantitatively using statistical methods of analysis. Analysis of the PIRLS data requires specific software that must be used when analysing the data. This study took cognisance of validity and reliability as they are crucial in research, hence they are outlined under the methodological norms section. Ethics are imperative in a research inquiry and the researcher abided by the ethical considerations discussed.

CHAPTER 5: DATA PRESENTATION AND RESULTS

5.1 INTRODUCTION

This chapter presents and discusses the results from the data analysis conducted to address the research questions. The data are presented in a three-fold format to answer the three sub-questions. Descriptive statistics were utilised to address sub-question 1. With the use of Rasch Measurement Theory (RMT), Rasch Analysis was conducted using the RUMM2030 software program to address sub-questions 2 and 3. Outputs from RUMM2030 presented in this study indicated items in the selected passage showing signs of measurement invariance. The items analysed were selected from the PIRLS 2021 passages, namely *The Amazing Octopus* passage. Section 5.2 centres on the South African PIRLS 2021 released results with a specific focus on the languages selected for this study. That was done to assist in addressing sub- question 1. Section 5.3 presents the findings and results from the Rasch Analysis and discussion thereof, to address sub-questions 2 and 3. Section 5.4 provides a summary of the items that depicted signs of Differential Item Function (DIF) in the two sampled languages before Section 5.5 concludes the chapter.

5.2 DESCRIPTIVE RESULTS TO ADDRESS SUB-RESEARCH QUESTION 1

It should be noted that this section is two-fold, Section 5.2.1 provides the overall South African PIRLS 2021 achievement scores of the Grade 4 learners who completed the test in either English or isiZulu. Thereafter, Section 5.2.2 provides descriptive statistics to address sub-question 1. For this study sub-question 1 was posed as follows:

1. *How do the overall Grade 4 English and isiZulu learners' reading literacy achievement scores differ on PIRLS 2021?*

5.2.1 PIRLS 2021 Overall Mean Scores

The PIRLS 2021 South African study was nationally representative based on nine provinces and 11 South African official languages (DBE, 2023); thus, generalisations can be made about South African Grade 4 learners' reading literacy. Since this study took the form of a secondary analysis of the PIRLS 2021 data, generalisations can be

made about the population of English and isiZulu learners. PIRLS 2021 sampled 12 426 Grade 4 learners across South Africa; however, for the purpose of this study, the focus was only on Grade 4 learners who completed the PIRLS 2021 assessment, specifically *The Amazing Octopus* passage, in English and isiZulu. The rationale for selecting these languages was that English learners are one of the highest performing learners in PIRLS trends (DBE, 2023; Motshekga, 2017) and of the African languages, isiZulu was the largest representative for Grade 4 learners (DBE, 2023) and it is one the most spoken language in South Africa (van Staden & Roux, 2022). Due to time constraints and resources, only one passage and its set of items were deemed suitable to address the research questions posed by this investigation, subsequently, sample reduction took place as this study chose to examine only one PIRLS 2021 passage and items. A total of 505 Grade 4 learners answered *The Amazing Octopus* passage and items in English and isiZulu. Table 5.1 illustrates the total number of learners who completed the items in English and isiZulu.

Table 5.1: Total number of learners who completed The Amazing Octopus passage and items in English and isiZulu

Passage Title	Total Passage Sample	English Learners	Isizulu Learners
The Amazing Octopus	505	226	279

In PIRLS, the achievement scale is 0 to 1000, with the average being 500 and a Standard Deviation (SD) of 100[12] (Yin et al., 2023; Howie et al., 2017). Based on the fact that education systems of the participating countries are different, PIRLS developed international benchmarks which build on each other, representing increasingly demanding reading comprehension skills and strategies with each higher benchmark (Howie et., 2017) — Low, Intermediate, High, and Advanced.

[12] A standard deviation of 100 was determined by the IEA.

- The *Low International Benchmark* is 400: learners who can read to locate and retrieve explicit information (DBE, 2023) (*cf.* Section 2.3.2 (processes of comprehension)
- The *Intermediate International Benchmark* is 475: learners begin to interpret and identify obvious reasons and give basic explanations based on the passage read (DBE, 2023) (*cf.* Section 2.3.2).
- The *High International Benchmark* is 550: learners are able to make intricate connections between events in the passage they are reading (DBE, 2023) (*cf.* Section 2.3.2).
- The *Advance International Benchmark* is 625: learners must integrate ideas as well as evidence across a text to appreciate the main themes of the passage (DBE, 2023) (*cf.* Section 2.3.2)

These benchmarks present information about what the learners can do at certain intervals of the scale.

South African Grade 4 learners who did not reach the Low International Benchmark were unable to retrieve basic information from the passages to answer to simplistic questions (DBE, 2023). The results shown in Figure 5.1 represents South African Grade 4 learner achievement scores per language (only the selected languages for this study) and the overall mean score attained by all the South African Grade 4 learners who completed the PIRLS 2021 achievement booklets.

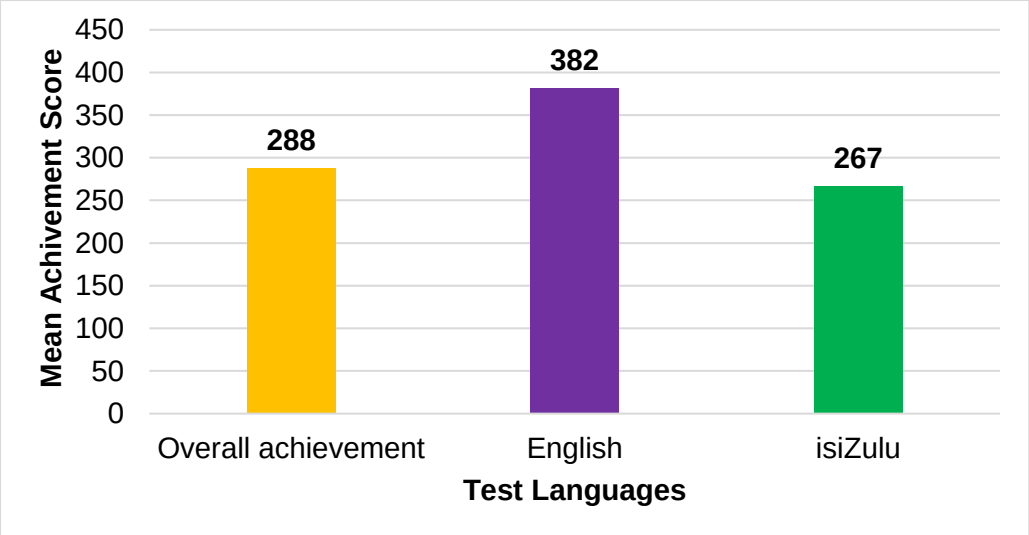


Figure 5.1: South African Grade 4 learner achievement scores in English and isiZulu

The overall mean score for Grade 4 learners tested in South Africa was 288 score points (SE=4.4)[13]. Learners who completed the test in English obtained 382 score points (SE=14.5) while learners who were assessed in isiZulu, obtained 267 score points (SE=6.5). It is clear that neither the two languages nor the overall South African achievement score reached the international mean score of 500. The learners who took the test in English were 18 points below the Low International Benchmark, whereas learners who took the test in isiZulu were 133 score points below. The results indicate that both groups of learners who completed the PIRLS 2021 assessments require more attention in terms of mastering the skill of reading literacy as neither of the two groups were able to reach the Low International Benchmark.

For the purpose of partially addressing sub-question 1, the means of the two languages were then tested to examine whether they are significantly different. There is a notable difference of 115 points between the two languages. It is interesting to note that the difference in the mean is significant, which means that the learners who were tested in isiZulu are approximately two years behind in terms of education when compared to those who took the test in English (*cf.* Howie et al., 2017; Section 1.3). This finding is consistent with what emerged during PIRLS Literacy 2016 (*cf.* Howie et al., 2017, Roux, 2020). This is of concern for isiZulu learners because they are in the same grade as the learners who took the test in English and have been at school for the same number of years. To illustrate this further, Table 5.2 shows the comparison of the mean scores of the two languages investigated.

Table 5.2: Language comparison of the Grade 4 learner achievement by language

Language	Mean	SE	English	isiZulu
English	382	14.5		▲
isiZulu	267	6.5	▼	

▲Significantly higher ▼ Significantly lower ● Not significantly different

[13] SE stands for Standard Error.

There is a significant difference in mean score between those learners who completed the PIRLS assessment in English when compared to those who completed it in isiZulu. The difference in mean score could possibly be due to several factors, such as teacher qualification, quality of translations, teacher content knowledge, parental involvement and learners' own attitudes towards reading. Further analysis was done on the raw scores of each language group to examine metric equivalence in the items of the passage.

5.2.2 Descriptive Statistics of Sub-Question 1

This section of the chapter deals with raw scores of each language group in an attempt to address sub-question 1. Differences in the mean scores as well as the percentage correct were also calculated. *The Amazing Octopus* had 15 items and the maximum score that a learner could obtain for this passage was 20. Table 5.3 depicts raw mean scores of the two languages.

Table 5.3: Raw mean scores per language

Passage	Overall Mean Score	English Mean Score	isiZulu Mean Score
The Amazing Octopus	7	10	2

The overall raw mean score attained by South African Grade 4 learners is 7. The isiZulu language group has the lowest mean of 2 whereas English language group has a raw mean score of 10 resulting in a difference of 8 between these two raw mean scores. It is clear that learners who completed the test in isiZulu found *The Amazing Octopus* passage and items more difficult than the English language group.

Table 5.4 depicts the number of the learners who correctly answered each of the 15 items as well as the percent correct.

Table 5.4: Number and percentage of learners who correctly answered items in English and isiZulu

Item No	English			isiZulu		
	N Completed	N Correct	% Correct	N Completed	N correct	% Correct
1	223	64	29	258	37	14
2	203	122	60*	232	65	28
3	178	70	39	221	57	26
4	190	94	49	211	81	37
5	196	37	19	230	4	2
6	203	97	48	226	10	4
7	202	122	60*	214	43	20
8	191	109	57*	185	38	21
9	208	49	24	198	3	2
10	185	67	36	173	3	2
11	191	83	43	173	37	21
12	185	51	28	168	5	3
13	170	85	50*	154	37	24
14	179	56	31	152	5	3
15	181	61	34	152	4	3

** Items correctly answered by 50% of the learners.*

Of the two languages, the English learners obtained the highest percentage of correctly responding to the items, with 60% of the learners answering items 2 and 7 correctly. Only 2% of isiZulu language learners were able to answer items 5, 9 and 10 correctly, which was the lowest percentage of learners answering an item correctly.

The table indicates that isiZulu learners who answered *The Amazing Octopus* assessment found it extremely difficult as none of the items were answered correctly by at least 50% of the learners. Whereas four of the items (items 2, 7, 8 and 13) were correctly answered by more than 50% of the learners who took the test in English. For isiZulu learners items 5, 6, 9, 10, 12, 13 and 14 are of concern because they are the items that have the lowest percentage of learners who correctly answered them. From the raw mean scores, *The Amazing Octopus* passage was found to be extremely

difficult for learners who completed the assessment in isiZulu and certain items, as presented in Table 5.4, proved challenging for those who took it in English.

From the analysis of the raw scores of those who responded to passage and items in English and isiZulu, it is thus evident that the items in the passage were non-equivalent as the item difficulty was not the same across the two languages. Due to the differences in the mean scores and raw mean scores of the two languages, a second set of analyses took place to further investigate where possible measurement invariance across the items could help explain the difference in the achievement scores of English and isiZulu language learners who responded to *The Amazing Octopus* items.

5.3 RASCH ANALYSIS EVIDENCE FOR SUB-QUESTIONS 2 AND 3

This section of the study discusses the results produced by RUMM2030 after the PIRLS 2021 data was inputted into the program to investigate for possible item bias. The current study investigated possible measurement of invariance/item bias and functionality[14] of the items in *The Amazing Octopus* passage answered by English and isiZulu language learners. As indicated, only the English and isiZulu language learners who were allocated *The Amazing Octopus* passage and its accompanying items were sampled, and their achievement data were inputted into the software. As this study utilised RUMM2030 to carry out Analysis of Invariance (ANOVA) statistics using Rasch Measurement Theory (RMT), the outputs provided this study with sufficient evidence to answer to sub-question 2 and 3. Sub-question 2 and 3 were posed as follows:

2. *To what extent can the difference in achievement be explained by possible bias (measurement invariance) between English and isiZulu responses during PIRLS 2021?*
3. *To what degree are items functionally equivalent post-translation?*

[14] Functional equivalence that has to do with instrument/item behaviour (*cf*. Section 3.7.2)

5.3.1 Summary Statistics for *The Amazing Octopus* Passage

This section of the study presents summary statistics (also known as reliability statistics) for *The Amazing Octopus* passage. Table 5.5 depicts summary statistics for the sample of this study, which includes English and isiZulu language groups.

Table 5.5: Summary statistics for The Amazing Octopus passage

Analysis name	Power of analysis of fit	PerSepIdx with extreme	PerSepIdx with no extreme	Coefficient Alpha with extremes	Coefficient Alpha with no extremes
Amazing Octopus	Good	0.82723	0.82874	0.91149	0.89583
English Only	Good	0.85259	0.84681	0.90977	0.89340
isiZulu Only	Too low	0.26960	0.22611	0.65675	0.59572

It is vital to note that if there is missing data, the Cronbach Alpha (α) cannot be calculated. During the calculation of summary statistics for *The Amazing Octopus* passage[15], and the two languages, the missing data were excluded, hence Table 5.5 shows the Cronbach Alpha. Furthermore, the acceptable α is .70 (Combrinck, 2020). The α for *English Only* is .90 which is excellent, and the power of analysis fit is good, signifying that the test items differentiated between different person abilities. In contrast, the α for *isiZulu Only* is .65 and it is below the acceptable value and the power of analysis is too low, signifying that the items in the test was not able to differentiate between the different person ability. According to Combrinck (2020), the minimum acceptable Person-Separation Index (*PerSepIdx*) is >0.7 which indicates statistically differentiable groups. Souza et al. (2017) states that the *PerSepIdx* indicates the acceptable level of separation within the persons. Furthermore, Souza

[15] The summary statistics calculated for *The Amazing Octopus* only included the sample for this study (505 learners sampled from English and isiZulu learners)

et al. (2017) highlighted that if the *PerSepIdx* values are less than the acceptable value, it implies that there is a need for items that will better separate the different abilities of persons. Table 5.5 indicates that the *PerSepIdx* for the passage *The Amazing Octopus*, was above the minimum acceptable level (0.82) thus implying that the groups were statistically differentiable, and the items differentiated between the two language groups compared. Similarly, for *English Only,* the *PerSepIdx* with no extreme is above the minimum level (0.84), thus the items were able to differentiate between high and low performers or person ability. However, for *isiZulu Only,* the *PerSepIdx* fell below the acceptable level (0.22), implying that the items did not differentiate between the high and low performers. As a result, item fit statistics were conducted to further look into the items.

5.3.2 Item Fit Statistics of *The Amazing Octopus* Passage

This section presents the item fit statistics for *The Amazing Octopus* passage. Rasch Analysis constructs a unidimensional scale and then tests how well the data fit that model through the notion of fit statistics that allows the determination of whether the items the respondent answered proceed in the order anticipated, based on the scale developed (Combrinck, 2020). The RMT assumes that the ability of the respondent increases when the items increase in difficulty (Bonne, 2016; Combrinck, 2018). The degree to which a parameter estimate is valid, is measured by statistics quantifying the fit of the data to the measurement model (Linacre, 1994). Item fit statistics are then used to determine the link between the persons and the items. However, if the results show a lack of fit, then the implication is that the person ability did not increase as the item difficulty increased. How well the data fits the Rasch model is indicated by the chi-square. According to Combrinck (2020), a significant chi-square illustrates that the data and the Rasch model are significantly different from one another, whereas a small chi-square which is not significant (p>0.05), is desirable and shows that the data fit the model. It is also vital to note that -2.5 to +2.5 is the ideal range that indicates item discrimination among the respondents. If the residual value is negative, it infers that the item was too easy (underfit) and when the residual value is positive it infers that the item was too difficult (overfit) (Combrinck, 2019).

To test the null hypothesis of this investigation, individual item fit statistics were utilised (Table 5.6). The null hypothesis of this investigation states that the mean scores of the learners who took the test in English and isiZulu are not different and therefore measurement invariance is not a contributing factor (H_0 = μEnglish = μisiZulu). If the null hypothesis is rejected, the alternative hypothesis states that the mean scores of the learners who took the PIRLS 2021 assessments are different and measurement invariance might be the contributing factor (H_a = μEnglish ≠ μisiZulu).

The results presented in Table 5.6 include the individual-fit statistics for an overall sample (n=505) of the English and isiZulu learners that completed *The Amazing Octopus* assessment.

Table 5.6: Individual Item-fit statistics for the Amazing Octopus Passage

Item	Difficulty	SE	Fit Residual	Chi-Square	Probability
Z03	-0,260	0,133	2,651*	12,88	0,025
Z04	-0,895	0,124	2,855*	14,44	0,013
Z08	-0,727	0,130	1,688	13,84	0,017
Z13	-0,517	0,145	2,170	2,89	0,717
Z02	-1,090	0,120	-1,262	12,10	0,034
Z05	1,978	0,199	-1,441	4,46	0,485
Z07	-0,965	0,124	-3,253*	33,62	0,000**
Z09	1,678	0,187	-2,911*	14,54	0,013
Z10	0,922	0,168	-3,503*	19,59	0,001
Z11	-0,275	0,140	-1,622	8,64	0,124
Z15	0,899	0,173	-2,286	9,56	0,089
Z01	-0,363	0,081	1,839	12,72	0,026
Z06	-0,540	0,083	-2,458	11,58	0,041
Z12	0,265	0,107	-2,057	13,47	0,019
Z14	-0,110	0,083	-0,525	10,86	0,054

* Fit residuals are shown if below -2.5 or above +2.5.

** Bonferroni adjustment is 0.000667 for all the items. Items that are smaller than the Bonferroni adjustment are highlighted, and they are significant.

Table 5.6 indicates that 5 of the 15 items in *The Amazing Octopus* passages showed misfit where the fit residuals were either above +2.5 or below -2.5. Items 3 and 4 displayed overfit where the items were too discriminating within the two language groups tested. On the other hand, items 7, 9 and 10 displayed underfit which implies that there was too little discrimination between the English and isiZulu language groups. In addition, item 7 displayed underfit that was significant, and it was the only item that displayed a significant misfit. The implication is there was too little discrimination in item 7 between the English language and isiZulu language groups. To further examine the null hypothesis of this investigation, it was crucial to look at the individual item fit statistics per language groups.

Table 5.7: Individual item-fit statistics for English language group

Item	Difficulty	SE	Fit Residuals	Chi-Square	Probability
Z03	0,272	0,199	1,444	9,64	0,008
Z04	-0,350	0,184	1,380	2,58	0,276
Z08	-0,822	0,183	1,723	13,75	0,001
Z13	-0,232	0,195	-0,158	4,31	0,116
Z02	-1,259	0,188	-0,594	1,57	0,456
Z05	1,881	0,226	-0,634	1,05	0,592
Z07	-1,284	0,187	-1,669	11,85	0,003
Z09	1,444	0,206	-1,979	8,94	0,011
Z10	0,537	0,194	-2,33	4,99	0,083
Z11	0,044	0,186	-1,141	4,27	0,118
Z15	0,679	0,199	-0,866	2,04	0,362
Z01	-0,047	0,119	1,757	3,84	0,147
Z06	-0,915	0,122	-1,42	3,06	0,217
Z12	0,123	0,134	-0,34	0,51	0,774
Z14	-0,072	0,103	0,406	3,71	0,157

* Fit residuals are shown if below -2.5 or above +2.5.

** Bonferroni adjustment is 000667 for all the items. All the items that are smaller than the Bonferroni adjustment are highlighted, and they are significant.

Table 5.7 shows the individual item fit statistics of the English language group only. The table indicates that none of the items in *The Amazing Octopus* passage were

either overfit (discriminating too high) or underfit (discriminating too little) within the English language group. None of the items in the English language group depicted misfit that was significant. It appears that for the English language group, all the items were a better fit.

Table 5.8 depicts the individual item statistics for the isiZulu language group.

Table 5.8: Individual item-fit statistics for isiZulu language group

Item	Difficulty	SE	Fit Residuals	Chi-Square	Probability
Z03	-0,919	0,167	1,875	1,837	0,399
Z04	-1,623	0,161	1,907	3,858	0,145
Z08	-0,759	0,193	0,822	5,014	0,081
Z13	-0,977	0,208	2,196	3,668	0,159
Z02	-1,066	0,163	-0,48	3,566	0,168
Z05	2,164	0,468	-0,776	0,507	0,776
Z07	-0,766	0,18	-1,623	6,956	0,030
Z09	2,586	0,628	-1,035	1,788	0,408
Z10	2,411	0,627	-0,908	1,773	0,412
Z11	-0,903	0,197	-2,864*	13,569	0,001
Z15	1,634	0,482	-1,011	0,555	0,757
Z01	-1,092	0,099	-2,366	5,169	0,075
Z06	-0,268	0,138	-1,064	2,554	0,278
Z12	0,126	0,197	-1,725	4,481	0,106
Z14	-0,547	0,133	-1,575	2,842	0,241

* Fit residuals are shown if below -2.5 or above +2.5.

** Bonferroni adjustment is 000667 for all the items. All the items that are smaller than the Bonferroni adjustment are highlighted, and they are significant.

Table 5.8 illustrates individual item statistics for isiZulu language group; only one item showed misfit. Item 11 depicted underfit, inferring that there was too little discrimination between the isiZulu languages learners that have higher and lower abilities.

Taking into account the individual item fit statistics of the overall sample (five items showing misfit; two overfit and two underfit and one significant underfit) and individual

item fit statistics per language group, the English group depicts no misfitting items within the group and the isiZulu group depicts one misfitting item that is underfit but not significant. It is clear that there is a degree of variation between the English and isiZulu language mean scores. Therefore, the above provides evidence to partly reject the null hypothesis.

5.3.3 Differential Item Functioning of Items in *The Amazing Octopus* Passage

This section provides a discussion of the items that displayed Differential Item Functioning (DIF) across the English and isiZulu language groups. When DIF occurs, it means that the persons having the same ability do not have the same chance of correctly responding to the item (Sandilands et al., 2013). DIF testing includes an ANOVA test that provides information on whether the mean scores of the groups examined are comparable or not. The ANOVA test conducted for this study includes a person factor analysis of all 15 items that accompanied *The Amazing Octopus* passage. The logic behind the use of an ANOVA test is twofold: firstly, to compare the mean scores of the two language groups investigated and secondly, to test the null hypothesis (Field et al., 2012).

The null hypothesis of this investigation states that the mean scores of the learners who took the test in English and isiZulu are not different and therefore measurement invariance is not a contributing factor ($H_0 = \mu$English $= \mu$isiZulu). If the null hypothesis is rejected, the alternative hypothesis states that the mean scores of the learners who took the PIRLS 2021 assessments are different and measurement invariance might be the contributing factor ($H_a = \mu$English $\neq \mu$isiZulu).

The ANOVA outputs provide the *F*-ratio and the *p*-value. The *F*-ratio is a measure where two quantities are expected to be approximately equal under the null hypothesis (Field et al., 2012). In the case of this study, the mean scores of the two language groups, respectively the English language and isiZulu language groups should have approximately the same mean scores that are not significantly different from each other. The *p*-value (probability value) is a statistical model that is conducted to test the null hypothesis in order to assign statistical significance of the evidence found (Cohen et al., 2018; Field et al., 2012). When put together, these values are able to help this

investigation test the null hypothesis and to inform this study whether the mean scores between the two language groups are equal. Table 5.9 displays the DIF summary for *The Amazing Octopus* passage, as indicated by ANOVA statistics.

Table 5.9: DIF Summary of The Amazing Octopus passage

Item	F-ratio	Probability
Z08	7,81919	0,005463
Z04	13,71932	0,000243*
Z08	2,07984	0,150234
Z13	4,97365	0,026550
Z02	2,88598	0,090232
Z05	1,62755	0,202886
Z07	14,55166	0,000158*
Z09	8,66164	0,003469
Z10	24,20394	0,000000*
Z11	3,07292	0,080613
Z15	7,10563	0,008137
Z01	9,52798	0,002163
Z06	47,59605	0,000000*
Z12	10,28951	0,001491
Z14	0,20134	0,654006

* Significant at 5 percent level (Bonferroni 0,001111)

It is important to note that if the *p*-value is less than <0.05, it presents robust evidence to reject the null hypothesis, whereas if a *p*-value is more than >0.05, it indicates weak evidence for rejecting the null hypothesis (Cohen, 2018; Field et al., 2012). In the case of this study, four of 15 items displayed *p*-values that enable this study to reject the null hypothesis, namely, items 4, 7, 10 and 6 and they showed signs of differential item functioning. The implication is that the four items, highlighted in Table 5.9, presents uniform DIF, inferring that these items functioned differently for English and isiZulu language groups. Andrich and Hagquist (2015) explained uniform DIF as a notion that presents consistent differences in the probabilities of persons with the same underlying abilities to correctly answer items when compared to different groups.

To further examine the DIF in *The Amazing Octopus* passage, Item Characteristic Curve (ICC) graphs were utilised. Figure 5.2 shows an example of an ICC graph.

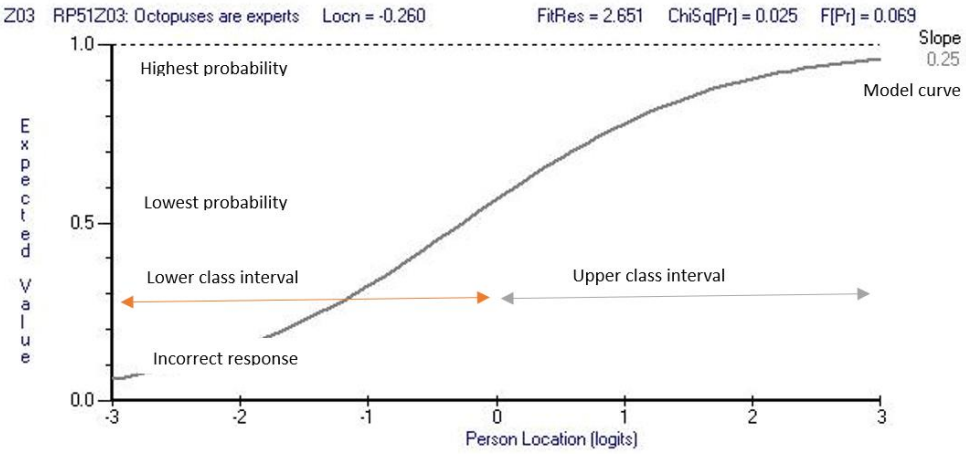


Figure 5.2: An Example of an ICC graph

The ICC graph has an *x* and a *y*-axis. The *x*-axis depicts person location which Combrinck (2020) explains as the respondents' abilities. The *y*-axis is the expected value commonly referred to as the probability score. At 0.0 where the x-axis meets the y-axis is incorrect response probability and the start of the lowest probability of a correct answer until 0.5, while 0.5 is the start of the highest probability of a correct response to 1.0 (Mtsatse, 2015). It is noteworthy to mention that at the *x*-axis, the ICC graph is divided into two interval classes namely, the lower-class interval and the upper-class interval. The lower-class ranges from -3 to 0 and the upper-class ranges from 0 to 3. The lower class is labelled as the persons with lower abilities to correctly answer to the items and the upper class is labelled as the persons with higher abilities to correctly answer to the items (Combrinck, 2020; Van Staden, 2018).

When the persons are placed on the person location (logits) scale ranging from -3 to 3, the assumption is that they have some sort of underlying ability. When the ICC of the group(s) investigated is below the model curve (usually represented in grey) it infers that the group found the item difficult. Each item presenting with DIF is represented below by the ICC graph, depicting differential item function of the specific item in both lower and/or upper-class interval. The ICC graphs provided evidence to answer sub- questions 2 (item difficulty/possible item bias) and 3 (item behaviour) of this inquiry, thus the two sub-questions are answered concurrently.

Item 4 in *The Amazing Octopus* passage

Item 4 in *The Amazing Octopus* passage is a Multiple Choice (MC) question that required learners to *Focus on and Retrieve Explicitly Stated Information*. The question was structured as follows:

4. Octopuses do not have bones. What does this mean they do?

a. Hide with other octopuses
b. Hold onto rocks
c. Fit into very tiny places* (correct answer)
d. Look like seaweed

Item 4 was phrased as follows in isiZulu:

4. Ama-okthophasi awanawo amathambo. Lokhu kusho ukuthi angakwazi ukwenzani?

a. Ukucasha namanye ama-okthophasi
b. Ukubambelela emadwaleni
c. Ukuzishutheka ezindaweni ezincane* (Impendulo elungile)
d. Ukubukeka njengokhula lwasolwandle

Figure 5.3 illustrates the ICC for item 4 of *The Amazing Octopus*.

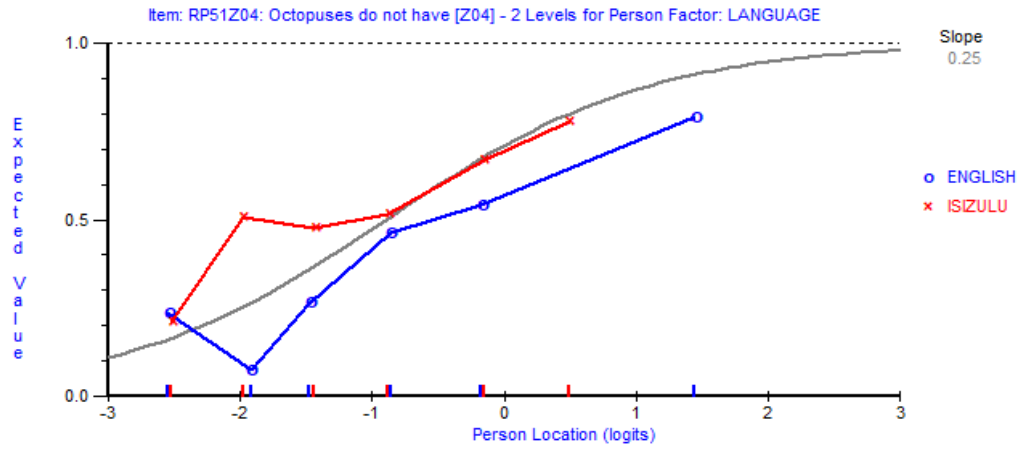


Figure 5.3: Item Characteristic Curve for The Amazing Octopus Item 4

For the English sub-group (represented in blue), it displays that those who were at the lower-class interval, specifically between -2 to -1.8, had the lowest probability (less than 10%) of choosing the correct answer and, as such, found the item very difficult. For the isiZulu sub-group (represented in red), those at the same lower-class interval, had approximately 50% change of correctly responding to the question. From the -0.8-person location onwards, the ICC changes in an upward direction for the English sub-group, meaning that the learners who are at the lower-class interval, have a slightly higher probability of choosing the correct answer, although they still found the item difficult as their ICC is below the model curve. Interestingly, the isiZulu sub-group, who are at the lower-class interval, found this item not too difficult as their ICC is above the model curve.

The English sub-group in the upper-class interval at 1.4-person location, have the highest probability (approximately 80%) of choosing the correct option, although their ICC is below the model curve. The isiZulu sub-group, who are at the upper-class interval at the 0.5-person location, also had a high probability (approximately 80%) of choosing the correct option. Overall, Item 4 functioned differently across the two language sub-groups and appears to discriminate against the English sub-group at both class intervals, inferring that item 4 was more difficult for the learners who wrote the test in English.

The next figure, Figure 5.4, depicts the distractor analysis for Item 4.

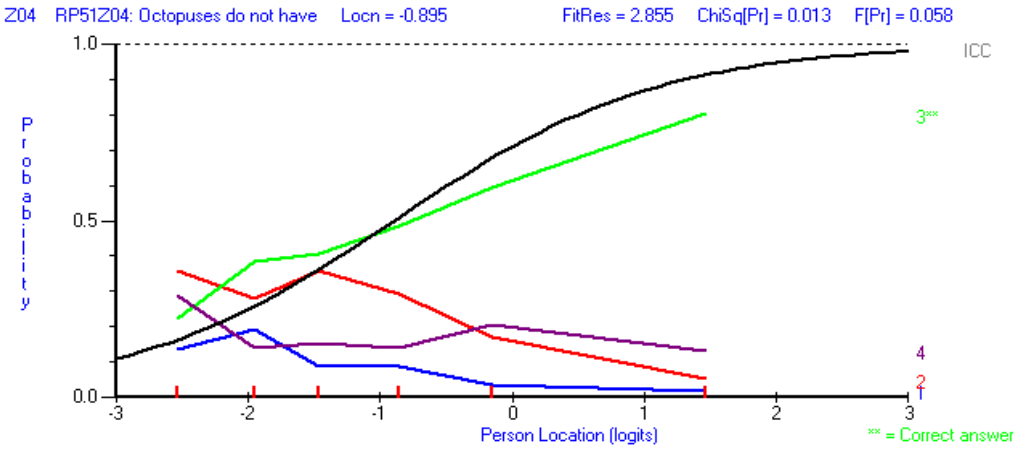


Figure 5.4: Distractor analysis in The Amazing Octopus for Item 4

For Item 4, the correct answer was distractor C (3). Between the -3- and -2-person location, learners were tempted to select distractor B (2) and distractor D (4). Learners had approximately a 20% chance of selecting distractor B (2). At the -1.9-person location, the learners had about a 40% chance of selecting distractor B (2). It would appear that from this point onwards, the learners were able to mostly correctly identify distractor C (3) as the correct answer. To illustrate this point, at the 1.5-person location, learners had about an 80% chance of correctly selecting distractor C (3). Next, is item 7 that displayed differential item functioning across the two languages.

Item 7 in *The Amazing Octopus* passage

Item 7 also displayed differential item function in the two languages investigated. Item 7 was a Constructed Response (CR) that required learners to *Make Straightforward Inferences*. It was phrased as follows:

7. What did Frieda the octopus learn to do? _____ (1 score point)

For isiZulu learners' item 7 was translated as follows:

7. Yini i-okthophasi u-Frieda eyafunda ukuyenza? _____ (iphuzu eliy-1)

Figure 5.5 displays the ICC for Item 7.

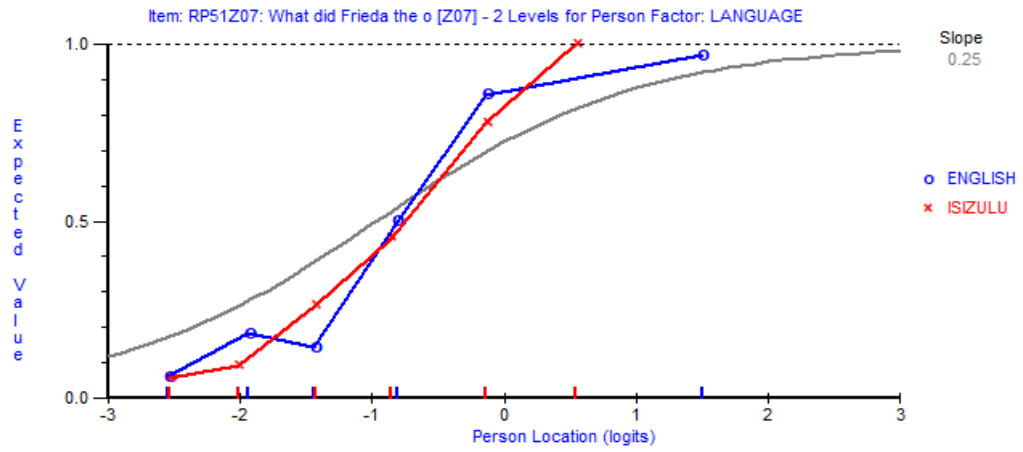


Figure 5.5: Item Characteristic Curve for The Amazing Octopus Item 7

Both English and isiZulu sub-groups, which were located at the lower-class interval from -3 to approximately -1.5, had the lowest probability of correctly responding to the question. Both English and isiZulu sub-groups, which were at the lower-class interval at -2.6-person location, had the lowest probability (less than 10%) of responding correctly to the item. Both sub-groups at the lower-class interval experienced inconsistency between the -3- and -1-person location. From the -0.8-person location onwards, the English sub-group experienced an upward direction whereby the sub-group had an approximately 50% chance of correctly responding to the question. The isiZulu sub-group experienced a gradual increase in probability of correctly responding to the item at -2-person location. At 0.6, the ICC shows that the isiZulu learners had the highest probability (100%) of responding correctly to the item and they may have found the item relatively easy as their ICC is above the model curve. The upper-class interval at 1.5 for the English sub-group also had a high probability (95%) of responding correctly to the question. Therefore, for this item, it is evident that it was inconsistent for both language groups, specifically at the lower-class interval. The question appears to be discriminating against both groups at the lower-class interval. Item 10 that displayed DIF is discussed below.

Item 10 in *The Amazing Octopus* passage

Item 10 displayed differential item function. This was a CR item that required learners to *Focus on and Retrieve Explicitly Stated Information*. The item asked:

10. Why do aquarium staff give octopuses puzzles?

_____ (1 score point)

For isiZulu learners' item 10 read as follows:

10. Kungani abasebenzi basethangini okuhlala kulo izilwanyana zasemanzini benika ama-okthophasi izinto zokuhlola ulwazi?

_____ (iphuzu eliy-1)

Figure 5.6 represents ICC graph for Item 10 that displayed differential item function.

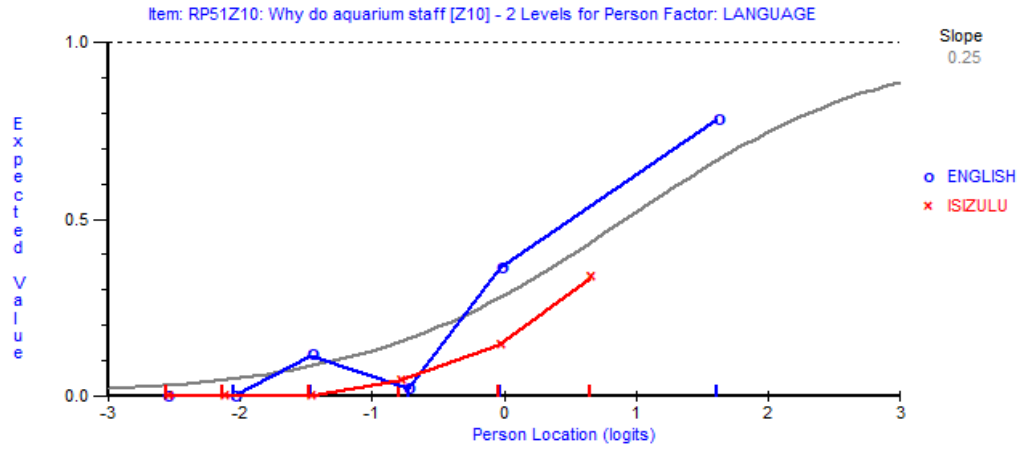


Figure 5.6: Item Characteristic Curve for The Amazing Octopus Item 10

Both English and isiZulu learners at the lower-class interval had the lowest probability of responding correctly to the question. Specifically at -2.5 to -1.5 logits, isiZulu learners had approximately 0% chance of responding correctly to the question. English sub-group at -2 logits also had the lowest probability and an approximately 0% chance of responding correctly to the item; however, change is seen at 1.5 and they are at approximately 10%, but this drops again to an approximately 0% chance. For English sub-group at approximately -0.7 to 1.6 logits, the sub-group experienced an upward direction whereby the sub-group had an approximately 80% chance of correctly responding to the question at 1.6-person location. For the isiZulu sub-group at -0.9-

person logits, the group had lowest probability and had less than 10% chance of correctly responding to the item. For both groups at the lower-class intervals, it appears that the item was too difficult. The isiZulu sub-group at the upper-class interval (approximately 0.7 logits) had the lowest probability of correctly responding to the questions (approximately 30%) whereas English sub-group (at approximately 1.8 logits), had more than 50% chance of responding correctly to the item. It is evident that this question had inconsistences, specifically in the upper-class interval. At the upper-class interval, the question appears to be discriminating against the isiZulu sub-group as their ICC is below the model curve. Item 6 that also displayed differential item function is discussed next.

Item 6 in *The Amazing Octopus* passage

Item 6 in *The Amazing Octopus* also displayed differential item function. This item is a CR item, worth two marks, that required learners to *Interpret and Integrate Ideas and Information* by asking the following question:

6. Give two ways octopuses escape their predators 1._____ (1 score point) 2._____ (1 score point)

For isiZulu learners' Item 6 was translated as follows:

6. Nikeza izindlela ezimbili ama-okthophasi asinda ngazo ezintweni eziwadlayo. 1._____(iphuzu eliy-1) 2._____(iphuzu eliy-1)

Figure 5.7 shows an ICC for Item 6 in *The Amazing Octopus* passage.

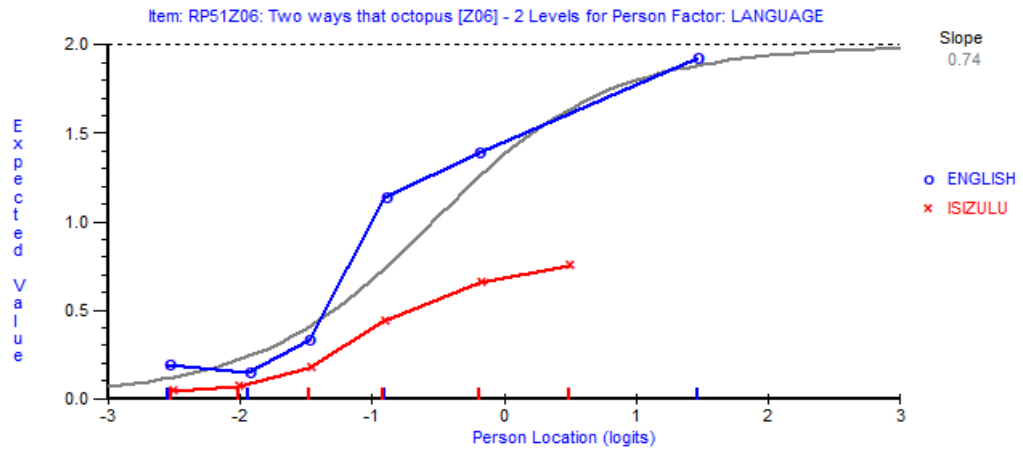


Figure 5.7: Item Characteristic Curve for The Amazing Octopus passage Item 6

To score two marks, learners had to provide two responses by listing them. If the learner only wrote one response instead of two, they would have been awarded 1 mark for partially[16] responding to the question. English and isiZulu groups who were located at the lower-class interval from -2.5 to 2 logits, had the lowest probability of correctly responding to the item (less than 20% chance of obtaining a partial mark). Furthermore, at -1.5 logits, the English sub-group experienced an upward direction in their ICC when compared to isiZulu, although with isiZulu language there was a gradual increase. At approximately -0.9 logits, the English sub-group had the highest probability (100%) of correctly responding to the question whereas, the isiZulu sub-group had the lowest probability (less than 50%) of correctly responding to the item and getting a partial mark at the same person location. In the upper-class interval at 0.5 logits, the isiZulu sub-group was leaning more towards the highest probability (approximately 70%) of partially responding to the item, but it appears that they found the item difficult. The English sub-group at the upper-class interval, had the highest probability of correctly responding to the question. It appears that this question discriminated against isiZulu learners at the upper-class interval as their ICC is below the model curve although they had a good chance of correctly responding to the item. It is also evident that this question was considerably more difficult for the isiZulu sub-

[16] Partial comprehension is worth 1 mark whereas, full comprehension is worth 2 marks.

group when compared to the English sub-group, as their ICC did not move above the model curve at any class-interval.

5.4 Summary of DIF items

Out of 15 items in *The Amazing Octopus* passage, four items displayed DIF across the two languages investigated in this study. A total of 27% of the items displayed differential item function. Two of the four items that behaved differently displayed discrimination against the isiZulu language sub-group but only one item discriminated against the English sub-group. Based on the DIF and ICC graphs for *The Amazing Octopus,* half the items that displayed DIF discriminated against the isiZulu sub-group with one item (Item 6), which was considerably more difficult for the isiZulu sub-group than for the English sub-group. Table 5.10 below depicts the summary of items that displayed DIF and the percentage of those items.

Table 5.10: Summary of items displaying DIF and percentage

Passage title	Total number of items	Total number of items displaying DIF	Percentage (%)
The Amazing Octopus	15	4	27%

5.5 CONCLUSION

In summary, Chapter 5 presented both descriptive and inferential statistics to address the research questions of this secondary analysis. Descriptive statistics mainly focused on the mean and raw scores of the two languages examined. They revealed that there is a large discrepancy between English and isiZulu language scores. Through significant testing of the mean of the two language sub-groups, it was evident that there is a significant difference between the mean scores of the two groups. The English sub-group performed significantly higher than the isiZulu language group (*cf.* Table 5.2). The analysis of the raw scores displayed in Table 5.4 revealed that the passage investigated was more difficult for learners who completed the assessment in isiZulu as none of items were correctly responded to by 50% of the learners.

However, four of the 15 items were scored correctly by 50% of the learners who took the test in English. The raw mean scores also revealed that English learners performed higher than those who responded to the passage and items in isiZulu (*cf.* Table 5.3). Thus, it is safe to infer that the passage items lacked metric equivalence.

Inferential statistics followed and focused on investigating whether any items in the PIRLS 2021 passage presented possible bias and thus, functioned differently for the two language groups. ANOVA statistics depicted that some of the items in the PIRLS 2021 *The Amazing Octopus* passage had different item difficulty for the two language sub-groups. Specifically, five of the 15 items displayed underfit and overfit, where two items displayed overfit (item 3 and 4) implying that they discriminated too much within the two language sub-groups. Items 7, 9 and 10 displayed underfit implying that they discriminated too little within the groups. For English learners all the items presented in Table 5.7 were a better fit as none displayed underfit or overfit. For the isiZulu language group, only one item showed underfit.

The Rasch Analysis revealed that four of the 15 items investigated displayed differential item function within the two language groups. Specifically, items 4, 7, 10 and 6 showed signs of DIF. It is important to note that they displayed signs of uniform DIF, implying that the persons (learners) of the same underlying ability did not have the same chance of correctly responding to the items and these items behaved differently for English and isiZulu sub-groups. The results provide this study with evidence to reject the null hypothesis that states that the mean scores of English and isiZulu learners are equal ($H_0 = \mu$English $= \mu$isiZulu). As the null hypothesis was rejected based on the evidence presented, the alternative hypothesis states that the mean scores of the learners who took the PIRLS 2021 assessments are different and measurement invariance might be the contributing factor ($H_a = \mu$English $\neq \mu$isiZulu). Subsequently the difference in the mean scores of the English and isiZulu sub-groups could be explained by the presence of measurement invariance/item bias in the items of *The Amazing Octopus* passage.

CHAPTER 6: CONCLUSIONS AND RECOMMENDATIONS

6.1 INTRODUCTION

This study examined the extent to which *The Amazing Octopus* items are equivalent across English and isiZulu language groups post-translation. Initially the measurement invariance was seen to be a factor in the difference in the achievement scores. As a result, this study made use of quantitative methods of investigating equivalence in *The Amazing Octopus* passage and its items post translation to ascertain whether measurement invariance was a contributing factor in the difference of the achievement scores. Through quantitative methods of research, this study was able to determine the extent to which items in *The Amazing Octopus* were equivalent. Since this study examined the difference in learner achievement, it took the form of a secondary analysis of PIRLS 2021 achievement data. During data analysis, descriptive analysis was conducted, calculating aspects such as raw scores, means and percentages to determine the overall difference in the scores of those who took the test in English and isiZulu. Thereafter, Rasch Analysis was conducted to determine whether items in *The Amazing Octopus* passage functioned differently in the selected languages, establishing functional equivalence.

The final chapter begins with a summary of the research findings presented for each of the sub-questions posed by this investigation (Section 6.2) and main conclusions are then drawn (Section 6.3). The following sections reflect on the methodology (Section 6.4) and this study's strengths and limitations (Section 6.5). Drawing on the results from this research, recommendations for policy, practice and research are offered in Section 6.6. Section 6.7 offers the potential contribution of this inquiry. Lastly, final thoughts bring this study to a conclusion.

6.2 SUMMARY OF THE RESEARCH FINDINGS

This study was motivated by South Africa's poor reading literacy achievement results emerging the PIRLS trend assessments (*cf*. DBE, 2023; Howie et al., 2006; 2011; 2017). Based on the PIRLS assessments, South African learners cannot read for meaning (DBE, 2023; Howie et al., 2017), implying that they have not developed the

skill of reading literacy, placing the country in a national 'reading crisis'. This means that the reading literacy levels remain low, and as a result, the Minister of Basic Education (DBE) (Motshekga, 2017; 2023) made it a national aim to raise the literacy levels. Literacy is an essential skill because the learner makes use of it on multiple occasions, such as for academic and social purposes. Specifically, this study stemmed from PIRLS 2021 as it revealed that there is discrepancy in the English and isiZulu achievement scores (DBE, 2023). Thus, the aim of this study was to determine whether possible measurement invariance could have resulted in the difference in the Grade 4 English and isiZulu learners' reading achievement scores during PIRLS 2021. In order to do so, this study explored metric and functional equivalence through the use of Rasch Analysis to determine whether any of *The Amazing Octopus* items functioned differently in the two languages.

Based on the above, this study investigated the following main research question: *To what extent are The Amazing Octopus items equivalent across English and isiZulu language groups post-translations?*

Three sub-questions were posed to help address the main research question and each is discussed in the following sections:

1. How do the overall Grade 4 English and isiZulu learners reading literacy achievement scores differ on PIRLS 2021?

In order to address the first sub-question, the overall results of the PIRLS 2021 assessment were considered. In addition, the raw scores and item percentage correct were considered across the two languages for the passage, *The Amazing Octopus*. This passage contained 15 items with a maximum mark allocation of 20.

The main finding for this sub-question was that the mean score of English and isiZulu language groups are significantly different. A 115-score point difference was evident between English (382 score points, SE=14.5) and isiZulu (267 score points, SE=6.5) language groups. To further investigate the mean score differences between the two languages, raw mean scores of *The Amazing Octopus* were also calculated. It was revealed that the English language group had a significantly higher score than the isiZulu language group. The learners who completed the passage in isiZulu found it

more difficult as none of the items were answered correctly by at least 50% of the learners. Four of the items were correctly answered by at least 50% of the English learners. The items that 50% of the English learners were able to correctly answer are categorised as the following processes of comprehension: *Focus On and Retrieve Explicitly Stated Information* (Items 2 and 8) and *Make Straightforward Inferences* (Items 7 and 13).

2. To what extent can the difference in the achievement be explained by possible bias (measurement of invariance) between English and isiZulu responses during PIRLS 2021?

After the initial exploration of the PIRLS 2021 achievement results across the two languages, Rasch Analysis was conducted in an attempt to further unpack the differences in mean score (possible bias) for those learners who completed the test in English or isiZulu.

Individual item-fit statistics were conducted to assist in partially accepting or rejecting the null hypothesis, $H_0 = \mu$English $= \mu$isiZulu. The item-fit statistics showed that items in *The Amazing Octopus* passage displayed misfit, where two of the items were underfit and three of the items were overfit. DIF was then conducted to highlight the differential functioning across English and isiZulu language groups. When the DIF was conducted, it included the ANOVA statistics which also helped in testing the null hypothesis. The ANOVA checked for significant differences between the English and isiZulu mean scores (p-value <0.05). The null hypothesis was rejected based on the findings of the overall mean scores, raw scores and ANOVA statistics. The following items indicated differential functioning: Items 4, 7, 10 and 6.

Based on the ANOVA statistics, it was evident that there was differential functioning of the items across the two languages, which led to an investigation of the Item Characteristic Curve (ICC) of the problematic items. Some of the items in the passage examined discriminated against the isiZulu language group when analysed against the English language group, although one item also discriminated against the English learners at a non-significant level.

In *The Amazing Octopus* passage, five of the 15 items lacked metric equivalence (item difficulty) as they demonstrated item discrimination among the learners of the different groups examined, which may explain the difference in the achievement scores of those who took the PIRLS 2021 assessment in either English or isiZulu.

3. To what degree are the items functionally equivalent post-translation?

Sub-question 3 was an extension of sub-question 2. The focus of this question was to further unpack the problematic items that displayed DIF. The examination of those items was conducted through the use of ICC graphs to showcase functional equivalence. Data from the RUMM2030 program revealed that four of the 15 items in *The Amazing Octopus* passage functioned differently for the two language groups. The items presented with uniform DIF which signifies that one language group had a consistently different probability of correctly responding to the items, despite the fact that the persons had the same underlying ability as a person who is in another group (Combrinck, 2020). For this study that meant that learners who took the test in isiZulu had a consistently different probability of correctly answering the item despite the fact that they are in the same lower and/or upper class as learners who took the test in English. The consistent difference in the probability may be due to translation problems. For example, two of the four items that displayed DIF revealed that the isiZulu sub-group had less chance of correctly responding to the question when compared to English sub-group, regardless of the fact that they were in the same class interval (*cf.* Figure 5.7, Item 6). As a result, four of the 15 items in *The Amazing Octopus* passage lacked functional equivalence post-translation because they displayed differential item function in English and isiZulu.

6.3 MAIN CONCLUSIONS OF THE STUDY

As the main research question of this inquiry aimed at determining whether possible measurement invariance could have resulted in the difference in the Grade 4 English and isiZulu learners' reading achievement scores during PIRLS 2021, it was evident from the analysis that the mean scores are different and some of the items lacked both metric and functional equivalence. Therefore, some of the items in *The Amazing*

Octopus passage were metric and functional non-equivalent. The following conclusions were systematically reached upon thorough examination of the items.

Main Conclusion 1: *Significant difference in mean scores across English and isiZulu language groups*

PIRLS 2021 results showed that South African Grade 4 learners performed poorly, and they did not reach the low international benchmark (DBE, 2023). However, emerging from the results, it was evident that the learners who completed the test in English (382 score points, SE=14.5) performed better than those who completed the test in isiZulu (267 score points, SE=6.5) and they are different at a significant level ($p<0.05$). Thus, this study further investigated the difference in the mean scores. Evidence for sub-question 1 illustrates that the raw mean scores for English and isiZulu languages are not similar as the mean score for English is 10 and for isiZulu is 2 with an average raw score of 7. The item percentage correct table revealed that at least four of the items in the passage were correctly answered by 50% of the English whereas none of the items were correctly answered by at least 50% of isiZulu learners.

Main Conclusion 2: *Lack of metric equivalence in The Amazing Octopus passage across English and isiZulu*

Peña (2007) explained that metric equivalence has to do with item difficulty across different cultures or groups. It was stated by Aegisdóttir et al. (2008) that if an instrument lacks metric equivalence, it threatens construct validity of the assessment. Evidence for sub-question 1, specifically raw score analysis (*cf.* Table 5.4), illustrated that items in *The Amazing Octopus* passage lacked metric equivalence. In all the items of the passage, none were correctly responded to by at least 50% of the isiZulu learners, whereas four of the items were correctly responded to by at least 50% of the English learners. The item fit statistics also revealed that some of the items were overfit and underfit, implying that the level of discrimination/item was not the same across the two groups. The different item difficulty of the items in the passage could be interpreted as having an association with the difference in the achievement scores. Subsequently, the presence of measurement invariance in the items helped this current investigation explain the difference in the learner achievement scores of those who took the test in English and isiZulu.

Main Conclusion 3: ***Evidence of functional equivalence in The Amazing Octopus across English and isiZulu***

Functional equivalence was explained by Aegisdóttir et al. (2008) as the behaviour of the items across different groups from different cultures. For this study, it meant the items behaving the same way for both English and isiZulu post-translation and adaptations. Evidence for sub-question 3 from the Rasch Analysis illustrates that four (Items 4, 6, 7 & 10) of the 15 items in *The Amazing Octopus* passage lacked functional equivalence. For instance, the ICC graph (Figure 5.6) illustrates that at the upper-class interval, the isiZulu sub-group at 0.8-person location had the lowest probability of correctly responding to the items (approximately 30%), whereas English sub-group at the same person location, had more than 50% chance of correctly answering the question. It was then evident that the item had inconsistencies and thus functioned differently for the two groups. The different item behaviour of the items in the two language groups threatens content validity of the instrument (Peña, 2007). The items examined behaved differently for the two groups therefore, it can be concluded that the items did not favour one group.

Main Conclusion 4: ***DIF shows that differences in item type matters***

PIRLS achievement booklets have different types of items, Multiple Choice (MC) items and Constructed Response (CR) items (Fishbein et al., 2023). *The Amazing Octopus* passage was mostly populated with the CR items (11 CR items) when compared to the MC items (4 MC items). From the analysis of the raw scores, it was evident that for isiZulu learners all the item types were problematic thus there is no specific pattern revealed by the data that suggest one item type was problematic when compared to the other. For English learners, there is also no specific pattern of problematic items. Two items where 50% of the learners were able to correctly answer, were MC items and the other two were CR items. However, the DIF analysis revealed that three of the four items that displayed DIF were the CR items with only one MC item displaying DIF. Furthermore, two of the items that displayed DIF (Items 6 and 10) were CR items; these two items displayed discrimination against the isiZulu sub-group. The different item behaviour of the two aforementioned CR items discriminating against isiZulu, could be understood as the isiZulu sub-group struggled to provide written responses to the questions.

6.4 METHODOLOGICAL REFLECTIONS

This study was quantitative in nature as it sought to investigate the difference in mean scores of English and isiZulu Grade 4 learners who responded to *The Amazing Octopus* passage items by examining equivalence of the items. As this study utilised the PIRLS 2021 achievement scores, it adopted a secondary analysis research design. Secondary analysis research design is concerned with the re-analysis of data that were collected for different purposes (McMillan & Schumacher, 2014). The main aim of this study was to determine the extent to which the items in the selected passage are equivalent across English and isiZulu. To ascertain whether there is a difference in the scores and the items are metric and functional equivalent, it was necessary to adopt a quantitative approach as it allowed the investigation of a phenomenon using numerical data (Mishra & Alok, 2017). After the IEA granted the researcher permission to access the English data files, they specified that permission request to access the isiZulu data files had to be sent to the country's National Research Coordinator (NRC) which is the South African Department of Basic Education (DBE). The availability of the translated instruments could be seen as a possible limitation. In order to obtain the translated instruments, the PIRLS NRC must give permission to use the translated instruments. However, this process proved to be a major hurdle of this investigation and as such, the original design of the study had to be changed to accommodate the length of time it took to gain permission. What could also be considered a limitation is that the researcher was not part of the original study (Zimmerman, 2010). However, the researcher was part of the scoring process for PIRLS 2021 as well as part of the reliability scoring and quality assuring process.

The investigation of *The Amazing Octopus* passage and items could have been further addressed by making use of mixed methods research whereby the quantitative aspects of the passage and items are research, followed by a more in-depth (qualitative) look of the passage translated into isiZulu. This extra step would require language experts to review the passage and critique it according to a set of guidelines. In doing so, it would provide a more holistic view of the PIRLS 2021 passage and possible reasons for isiZulu learners' poor achievement in comparison to the English learners' achievement.

6.5 STRENGTHS AND LIMITATIONS OF THE STUDY

This current study was structured to investigate possible measurement invariance in PIRLS 2021 achievement booklets to determine whether the items were equivalent across the languages. It was formulated based on the poor South African reading literacy results emerging from the PIRLS 2021 assessment. To determine whether measurement invariance is a factor, item bias and equivalence were investigated in the English and isiZulu national versions of the instruments. The instruments were methodologically examined using one passage, namely *The Amazing Octopus*. As quantitative methods allow the researcher to study the phenomenon, provide robust view and extensive answering of the research questions (Pham, 2018), quantitative methods were deemed fit for this study because it allowed the researcher to use the PIRLS 2021 numerical datasets to address the research questions posed by this study and provide a clear picture of the phenomenon studied.

The PIRLS 2021 datasets were used to address the sub-questions. The PIRLS datasets used in this inquiry were re-coded before the analysis could start. The datasets were prepared through a series of steps (Combrinck, 2019) specifically for RUMM2030 and to fit the objectives of this inquiry. This study began by determining whether there is a significant difference in the achievement scores of the two languages by looking at the mean scores and raw mean scores (sub-question 1). Due to the findings in sub-question 1, this study then followed up with sub-question 2 that investigated whether the significant difference found could be explained by possible measurement invariance. This question specifically dealt with ascertaining metric equivalence in the items of the selected passage (item difficulty). This study further looked into instrument behaviour in the two mentioned languages to determine the behaviour of the items (sub-question 3). To address the research questions raw scores, percentages, mean scores (using IDB-Analyzer plug-in with SPSS) were calculated, and Rasch Analysis (using RUMM2030) was conducted to enable this study to reach the main conclusions presented. Subsequently, the conclusions reached by this inquiry were systematically generated based on the results and findings, the potential contribution this study might have, and recommendations made by this inquiry were informed by the findings from the research questions.

The limitations to this study are that this inquiry was structured to use a mono-method, discarding the valuable information that may have been yielded from a mixed methods study. Secondly, this study did not further explore the underlying factors that might explain the differential item function found in the items of the passage examined. For example, this study could have shared valuable insights if language experts such as translators and Grade 4 language teachers were consulted about the items that showed differential item function. As noted above, a PIRLS achievement booklet is made up of two passages and set of items; however, for this study only one informational passage was systematically examined to determine the difference in the achievement and ascertain equivalence in the items, excluding the second passage (Literary passage) in the PIRLS achievement booklets. Lastly, the sample size was reduced from 12 426 learners (nationally representative sample) to 505 due to the fact that this study sampled only English and isiZulu learners that responded to *The Amazing Octopus passage*. The reduced sample size means that the findings from this study are not generalisable to the Grade 4 South African population.

6.6 RECOMMENDATIONS FOR POLICY, PRACTICE AND RESEARCH

Based on the results founds in during this study, the following recommendations are made for policy (6.6.1), practice (6.6.2) and research (6.6.3).

6.6.1 Recommendations for Policy

South African schools should consider exposing their learners to informational texts, not only narrative texts. Exposure might prove beneficial since learners will have a greater familiarity with the texts. By reading a variety of texts, it may boost their vocabulary and their skill in reading such text. Moreover, it will be advantageous for the learners to practise with different texts as it may enrich their understanding of the world, learn new things and new words and/or phrases. It was revealed by the current research that Grade 4 isiZulu learners performed poorly when compared to those who answered the items of *The Amazing Octopus* in English. However, even though the English learners performed better, they also did not reach the low international benchmark. Therefore, it is recommended by this study that South African Grade 4 learners be continuously given access to a wide range and variety of material,

especially in African languages. The implication is that there should be policy, transcribed in the South African curriculum that stipulates that challenging materials should be integrated and used in the classroom specifically for reading, working towards integrating diverse texts and longer texts into learners daily reading. As noted in Section 2.3.2, the PIRLS reading comprehension processes can be used as guidelines for the teaching of reading literacy in the Foundation Phase.

Cross cultural instruments require extensive translation, but for South Africa that is proving difficult because of the unavailability of words and/or phrases and agreement among dialects in African languages (Howie et al., 2017; van Staden, 2006). Because of the complexity added by the diverse non-homogenous language context of South Africa, a series of language experts should be consulted and involved in the translation of instruments including the Pan South African Language Board (PanSLAB). Additionally, this study recommends that the issues of different dialects in each of the African languages be given attention (Mtsatse, 2017; Section 1.1.1), because learners of the same language have different dialects depending on districts and regions in which they live. The role of language in education (assessments) should be realised and treated as important.

Due to language differences and oppression of the African languages pre-1994, South Africa introduced the LiEP to ensure that indigenous languages were recognised and used in the education system. However, the South African DBE should consider the importance of a carefully planned implementation of the LiEP because South African classrooms are comprised with learners from diverse backgrounds with different home languages thus making it difficult to learn and take assessments in the language that is not truly their home language. PIRLS 2021 revealed that 70% of learners indicated that they spoke the language of test at home while 30% of learners indicated that they do not speak the language of test at home. Even though 70% of the population indicated that they speak the language of test at home, the remaining 30% who do not, indicates that it is crucial when looking at the differences in the scores. Perhaps this issue might also explain is the reason for a significant difference of the scores in international comparison (DBE, 2023). As noted in Section 3.2.1, the LiEP policy is there but implementation is questionable (Nugraha, 2019). It is concerning that Grade 4 learners cannot read for meaning in any language. Therefore, this study

recommends that the DBE should review the LiEP to ensure effective practice of the policy.

6.6.2 Recommendations for Practice

Teacher training is required to extensively equip teachers with the necessary skills on how to teach reading literacy and prepare them on how to adequately handle and integrate informational texts in the classroom. A series of workshops and training are a necessity. Once the teachers are trained with the necessary skills, knowledge and competencies, they will be able to equip children with the necessary skills to read for meaning because they will be guided by policy that is informed by research. It is crucial to mention that the DBE in collaboration with the South African Council for Educators (SACE), have a Continuing Professional Teacher Development (CPTD) system in place to encourage teachers to take part in formal and informal professional development activities to enhance their teaching and content knowledge (van Staden & Roux, 2022).

All learners need strong reading instruction regardless of whether they are struggling or not, to shift from non-reading to proficiency reading, learners should be taught phonics for decoding and spelling for encoding words (Herron & Gills, 2020). Teachers are encouraged to teach reading on an everyday basis while incorporating the five components of reading envisaged by the education system (DBE, 2011). To teach learners how to read, Gentry (2023) states that explicit and systematic teaching enables the learner to learn how to read. Thus, it is recommended that the teachers develop an explicit and systematic way of teaching reading to enable the learner to learn how to read for meaning. To afford teachers the opportunity to develop learners reading skills, it is recommended that teachers attend seminars and any developmental opportunities aimed at developing their abilities and understanding of teaching reading in a systematic method (Singh & Shaari, 2019).

Teachers must be well informed about the LiEP policy and what it means for classroom practice. Teachers need to be trained to fully comprehend the role of language in education because they are presented with learners from diverse backgrounds and the linguistic difference that learners present in the classroom requires teachers to

accommodate them. Izevbigie (2021) proposed that in the multilingual South African context, code-switching[17] and translanguaging[18] should be utilised as a strategy to differentiate curriculum and assessments to enhance success for learners in Foundation Phase classrooms. The recommendation for practice is that teachers should perhaps adopt translanguaging where possible in an attempt to cater for different kinds of learners speaking different HLs. It is also crucial to note that language switching in the class may be difficult as it may impede learner understanding during the process of teaching and learning when the teacher switches between languages. It is commendable that the DBE introduced the Integrated Quality Management System (IQMS) that highlights the gaps in teachers; professional performance and ultimately offer ways in which teachers can develop professionally. As teachers in a South African context are required to work in multilingual and multicultural classrooms, it is crucial that they take part in IQMS (van Staden & Roux, 2022) to improve their teaching (i.e., Strategies and methods) and to enhance their content knowledge.

6.6.3 Recommendations for further research

The following recommendations are made based on the findings of this study and the aim of this inquiry.

1. This study aimed at investigating the difference in learner achievement of those who took the PIRLS assessments in English and isiZulu, therefore, this current study recommends that investigations be launched into other South African languages that are not yet examined.
2. Furthermore, this research only examined one PIRLS informational passage, therefore, it recommends that more PIRLS passages be investigated to examine possible differential item function across the different South African languages.

[17] Code switching is when more than one language in the classroom is used as an approach to teaching of the lesson to accommodate learners speaking different kinds of languages.

[18] Translanguaging is the ability to move fluidly between languages and a pedagogical approach to teaching in which teachers support this ability.

3. Research should be conducted on how to effectively translate cross-cultural instruments in the South African context to attain optimal assessment validity and equivalence while taking into cognisance the different languages used in South Africa.
4. Research can be conducted on how to better accommodate the diverse linguistic issues with which South Africa is presented and how assessments can be differentiated to accommodate learners speaking different languages in one classroom, because current research indicated that 30% of learners indicated that the language of the PIRLS test is not the language they speak at home.
5. Lastly, this research recommends that extensive research be launched on how to refine the South African curriculum to meet the international standards. Curriculum refinements will also be advantageous in developing the reading literacy skills of South African learners in order to better prepare them for tertiary studies and the world of work. In short South African needs a functioning curriculum that will better prepare learners to meet the international standards.

6.7 POTENTIAL CONTRIBUTION OF THIS INQUIRY

After a methodological examination of *The Amazing Octopus* passage, this study was able to determine the extent to which the items were equivalent across two languages. Firstly, this study started by calculating the difference in the learner achievement scores in English and isiZulu to determine whether there is a difference and whether the difference is significant. Then, the calculation of raw scores were finalised. Thereafter, this study conducted a Rasch Analysis to attempt to explain the difference in the achievement scores and provide a more renounced understanding of the PIRLS 2021 results in the passage selected. The reason was to assess whether the difference in the achievement scores can be explained by possible measurement invariance. Thereafter, this study examined instrument behaviour in the two languages to examine whether the instrument items behaved differently, through the use of RUMM2030 outputs. As this study examined an ILSA passage, it has the potential of offering a valuable contribution to the ILSA literature, specifically for equivalence of ILSA instruments across languages.

Roux (2020) explained that ILSAs involve the testing of participants across diverse cultural groups and languages, therefore, it is crucial to note the difference in the languages by investigating different types of equivalence of the ILSA instruments. As such, this current study attempted to examine metric and functional equivalence in an ILSA instrument, to determine whether the instrument has the same item difficulty post-translation and behaves the same across two groups (English and isiZulu). Based on the findings of this study and a thorough review of the literature, this study contributes to assessment literature by revealing that more focus and effort should be put into the translations of cross-cultural instruments such as the PIRLS achievement booklets to ensure equivalence.

As the research design of this study permits the researcher to re-analyse the already collected data for the purpose of studying what remains to be learnt about the primary data (Johnston, 2014; McMillan & Schumacher, 2014), this study was able to ascertain whether the measurement invariance is a factor that might have a contribution in the difference in the learner achievement. It was already known that there is a difference in the learner achievement, however, because of this current inquiry, it was further learnt that the lack of equivalence in some of the items might have a contribution in the difference of learner achievement scores, specifically in the English and isiZulu language groups. This study has the potential of prompting researchers to investigate DIF in the remaining South African languages that were not examined in this study, to ascertain whether measurement invariance could be a contributing factor in the learner achievement scores of the languages not included in this inquiry.

PIRLS 2021 results indicated that the highest performing languages are English and Afrikaans with the indigenous lagging behind; however, none of the languages were able to reach the low international benchmark of 400 score points (DBE, 2023). It is concerning that there is a gap in learner achievement scores across all national languages, subsequently, that presents an opportunity for research to be done in investigating possible factors that might explain the gap in the learner's achievement scores. This study was constructed to investigate such; however, only two languages were considered. This study has the potential to contribute to the assessment literature by identifying that there is a significant gap in the learners scores and measurement invariance could help explain the gap in the reading scores.

6.8 FINAL THOUGHTS

Pre-1994, South Africa education provision was drawn along racial lines. Since the dawn of democracy, the education system has undergone transformations that introduced new policies and curricular reforms to offer equal access to all South African children. LiEP was crafted to recognise the different languages present in South Africa and introduce African languages into the curriculum. After a number of curricula reforms, CAPS was introduced in South African public schools to guide teaching and learning in schools. Despite the attempts made in the South African education system and curriculum, South African learners have performed poorly in PIRLS trend assessments (DBE, 2023; Howie et al., 2006, 2011, 2017).

This study was then aimed at determining whether possible measurement invariance could have resulted in the difference in the Grade 4 English and isiZulu learners' reading achievement scores during PIRLS 2021. It examined the difference of the learner achievement and the different types of translation equivalence since the PIRLS international instruments were translated to the South African languages in accordance with the LiEP. This study found that there is a significant difference in the learner's achievement scores of those who completed the PIRLS assessments in English and isiZulu, and the significant difference can be explained by possible measurement of invariance and varying instrument behaviour.

The main purpose of administering equivalent and valid assessment is to place all the learners taking the assessment on the same scale and thus obtain reliable and valid results. If the items that shows signs of DIF continue to be used to assess learners, assessment equivalence and validity are threatened, thus rendering the results not reliable. If South Africa continues to participate in International Large-Scale Assessments, innovative approaches to translation that are sympathetic to languages whose vocabulary has not yet evolved to incorporate modern and scientific technological vocabulary should be sought with the aim of crafting equivalence assessments. Prah (2006) claimed that African language-speaking majorities in South Africa need to ensure that in a short space of time African languages be developed into languages of science and technology and become equals, culturally and

linguistically with English-speaking and Afrikaans-speaking communities in South Africa.

In addition, education system transformation is a prerequisite as is curriculum reform in order to meet the international standards and raise reading literacy levels in South Africa. This begins at home with parental involvement and continues at Foundation Phase level where learners are taught explicit reading literacy skills on a daily basis with time being given to develop a culture of reading through access to a variety of reading material and resources. To respect and fulfil the right to basic education, a child must be able to read and write with understanding in their Home Language (HL) at a basic level (Spaull, 2021).

"Funda Mntanami"

Mrs B.N Mthimkhulu

LIST OF REFERENCES

Aegisdóttir, S., Gerstein, H. L., Cinnabars, D. C. (2008). Methodological Issues in Cross-Cultural Counselling Research: Equivalence, Bias, and Translations. *Division of Counselling Psychology*, *36*(2), 188-200. https://doi.org/DOI:10.1177/0011000007305384.

Affrman, I. (2010). Equivalence of Translations in International Reading Literacy Studies. *Scandinavian Journal of Educational Research*, *54*(1), 37-41. https://doi.org/DOI:10.1080/00313830903488460.

Alavi, M.S., & Bordbar, S. (2017). Differential Item Functioning Analysis of High-stakes Test in Terms of Gender: A Rasch Model Approach. *Malaysian Online Journal of Education Science, 5*(1), 10-12. https://files.eric.ed.gov/fulltext/EJ1125062.pdf

Alharbi, A. A. (2017). *The Translation and Translation verification of the PIRLS Reading Questionnaires for Saudi Students.* [Masters Dissertation]. University of Kansas. https://doi.org//www.10.5281/zenodo.887089.

Alomari, T. (n.d). Fundamentals of Social Science Research. *Academia*. 1-3. https://www.academia.edu/39098355/Fundamentals_of_Social_Research_Method

Andersson, A., & Osterling, L. (2014). Metric Equivalence in International Surveys: Cultural Edges. In Nicol, C., Liljedahl, P., Oesterle, S., & Allan, D. (Eds.), *Stockholm University, 2*, 2-5.

Andrich, D., & Hagquist, C. (2015). Real and Artificial Differential Item Functioning in Polytomous Items. *SAGE Journals, 75*(2), 1-5. https://doi-org.uplib.idm.oclc.org/10.1177/0013164414534258.

Apuke, O. (2017). Quantitative Research Methods: A Synopsis Approach. Arabian *Journal of Business and Management Review, 6*(11), 40-47. https://doi-org.10.12816/0040336.

Barbe, K. (1996). The Dichotomy Free and Literal Translation. *Northern Illinois University*.

Basterra, M. R., Trumbull, E., Salono-Flores, G. (2011). *Cultural Validity in Assessment.* (1st ed.). Routledge publishers.

Behr, D. (2017). Assessing the use of back translation: the shortcomings of back translations as a quality testing method. *International Journal of Social Research Methodology*, *20*(6), 573-584. https://doi.org/10.1080/13645579.2016.1252188.

Berman, A. I., Haertel, E. H., & Pellegrino, J. W. (2020). Introduction: Framing the issues. In A. I. Berman, E. H, Haertel & J. W. Pellegrino, (Eds.), *Comparability of large-scale educational assessments: Issues and recommendations.* (pp. 1-8) National Academy of Education. https://doi.org/https://naeducation.org/wpcontent/uploads/2020/05/Comparability-of-Large-Scale-Educational-Assessments_final.pdf.

Beukes, A. (2009). Language Policy Incongruity and African Languages in Post-apartheid South Africa. *Language Matters, 40*(1), p.35–55. https://doi.org/10.1080/10228190903055550.

Bonne, J. W. (2016). Rasch Analysis for Instrument Development: Why, When, and How? *CBE Life Sciences Education*, *4*, 1-7. https://doi.org//doi:10.1187/cbe.16-04-0148.

Brown, J. D. (2000). What is construct validity? *Shiken: JALT Testing and Evaluation, 4*(2), 9-12. https://hosted.jalt.org/test/PDF/Brown8.pdf.

Byrne, B. M., & Watkins, D. (2003). The issue of measurement invariance revisited. *Journal of cross-cultural psychology, 34*(2), 155-160. Doi:10.1177/0022022102250225.

Chan, D. N. S., & So, W. K.W. (2017). Translation and Validation of Translation in cross-cultural research: Strategies used in a study of cervical cancer screening

among ethnic minorities. *International Journal of Nursing Practice, 23,* 1-6. https://doi.org/10.1111/ijn.12581

Chen, C. (2010). On Reading Test and its Validity. *Asian Social Science, 6*(12), 191-192. Doi:10.5539/assv6n12p192

Cheung, G. W., & Rensvold, R. B. (2002). Evaluating the goodness-of-fit indexes for testing measurement invariance. *Structural Equation Modelling, 9*(2), 233-245. Doi:10.1207/S15328007SEM0902_5.

Cockle, M. (2023). Creating the PIRLS 2021 international database. In M. von Davier, I. V. S. Mullis, B. Fishbein, & P. Foy (Eds.), *Methods and Procedures: PIRLS 2021 Technical Report* (pp. 7.1-7.16). Boston College, TIMSS & PIRLS International Study Center.

Coetzee-Van Rooy, S. (2018). Dominant Language Constellations in Multilingual Repertoires: Implications for Language-in-Education Policy and Practices in South Africa, *Language Matters, 49*(3), 19-20, https://doi.org/10.1080/10228195.2018.1493613.

Cohen, L., Manion, L., Morrison, K. (2011). *Research Methods in Education*. (7th ed.) Routledge Publishers.

Cohen, L., Manion, L., Morrison, K. (2018). *Research Methods in Education*. (8th ed.) Routledge Publishers.

Combrinck, C. (2018). *The use of Rasch Measurement Theory to Address measurements and analysis Challenges in social science research.* [Doctoral Thesis]. University of Pretoria.

Combrinck, C. (2019). *Rasch unidimensional measurements models: Analysis data in RUMM2030 handout*. Centre for Evaluation and Assessment.

Combrinck, C. (2020). Is this a useful instrument? An Introduction to Rasch Measurement Models. In S. Kramer, S. Laher, A. Fynn, & H. H. Janse van

Vuuren (Eds.), (pp. 1-48), *Online Readings in Research Methods*. Psychological Society of South Africa: Johannesburg. https://doi.org/10.17605/OSF.I0/BNPFS

Conrad, C. F., & Serlin, R. C. (2011). *The SAGE Handbook for Research in Education* (2nd ed.). SAGE Publishers.

Costa, C., & Zezlina. E. (2013). The "Little readers' circle": *infectious love for reading in a middle school classroom*. IGI Global publishers.

Creswell, J. W., & Plano Clark, V. L. (2007). *Designing and conducting mixed methods research.* SAGE Publishers.

Creswell, J. W., & Plano Clark, V. L. (2018). *Designing and Conducting Mixed Methods Research.* (3rd. ed.). SAGE Publishers.

De Bruin, I. (2010). *Exploring how objects used in a Picture Vocabulary Test influence validity*. [Master's dissertation]. University of Pretoria.

Department of Basic Education (DBE). (2004). *National Norms and Standards for School Funding.* Pretoria Government Publishers.

Department of Basic Education (DBE). (2011). *Curriculum and Assessment Policy Statement. Grades 4-6. English Home Language*. Government Publishers.

Department of Basic Education (DBE). (2014). *Report on the annual national assessment of 2014: Grades 1–6 and 9.* Government Publishers.

Department of Basic Education (DBE). (2017). *The SACMEQ IV project in South Africa: A study of the conditions of schooling and the quality of education*. Government Publishers.

Department of Basic Education (DBE). (2023). *PIRLS 2021: South African Main Report*. Department of Basic Education: Pretoria.

Department of Education (DoE). (1997). *Language in Education Policy*. Pretoria Government Publishers.

Department of Education – Louisiana Believes. (n.d). *Guide for determining text complexity: Kindergarten through Grade 12.* https://education.illinoisstate.edu/downloads/casei/guide---how-to-determine-text-complexity-grades-k-12.pdf

Diaz, E., Brooks, G., & Johanson, G. (2021). Detecting Differential Item Functioning: Item Response Theory Methods Versus the Mantel-Haenszel Procedure. *International Journal of Assessment Tools in Education*, *8* (2), 376-378 https://doi.org/10.21449/ijate.730141.

Ebbs, D., Flicop, S., Hidalgo, M. M., & Netten, A. (2023). Systems and instrument verification in PIRLS 2021. In M. von Davier, I. V. S. Mullis, B. Fishbein, & P. Foy (Eds.), *Methods and Procedures: PIRLS 2021 Technical Report* (pp. 5.1-5.24). Boston College, TIMSS & PIRLS International Study Center.

Ebbs, D., & Wry, E. (2017). *Translation and Layout Verification for PIRLS 2016*. Boston College, TIMSS & PIRLS International Study Center.

Izevbigie, E. (2021). *Perceptions of foundation phase mainstream schoolteachers regarding differentiated assessment*. [Master's Dissertation]. University of Pretoria.

Ehsani, F., Zohrabi, K. (2014). Assessing Translation of Advertising Text (English Persian) Based on House's Functional-pragmatic Model of TQA. *Social and Behavioural Sciences 98*, 420-424. https://doi.org/doi:10.1016/j.sbspro.2014.03.435.

Field, A., Mile, J., & Field, Z. (2012). *Discovering statistics using R*. SAGE publishers.

Fishbein, B., Yin, L., & Foy, P. (2023). *PIRLS 2021 User Guide for International Database.* Boston College, TIMSS & PIRLS International Study Center.

Fleisch, B. D. (2008). *Primary Education in Crisis: Why South African Children Underachieve in Reading and Maths.* Juta Publishers.

Foy, P. (2018). *PIRLS 2016 user guide for the international database.* Boston College, TIMSS & PIRLS International study Center. https://TIMSSandpirls.bc.edu/pirls2016/internation-database/downloads/P16UserGuide.pdf

Fraenkel, R. J., Wallen, E. N., & Hyun, H. H. (2019). *How to design and Evaluate Research in Education* (10th ed.). McGraw Hill Education

Frankel, K. K., Becker, L. C. B., Rowe, W. M., & Pearson, D. P. (2016). From "What is reading?" to What is Literacy? *Journal of Education*, *196* (3), 6-7. https://doi.org/10.1177/002205741619600303.

French, B. F. (2014). Test Bias. In: A.C Michalos. (Eds). *Encyclopaedia of Quality of Life and Well-Being Researc*h. Springer Publishers. https://doi.org/10.1007/978-94-007-0753-5_2998.

Gareis, R., & Grant, W. L. (2015). *Teacher-made Assessments: How to Connect Curriculum, Instruction and Student Learning.* (2nd ed.). Routledge.

Gentry, J. R. (2023, September 20). Elite Universities call for change in reading education. https://www.psychologytoday.com/us/blog/raising-readers-writers-and-spellers/202309/elite-universities-call-for-change-in-reading?eml

Govender, R., & Hugo, A. J. (2020). An analysis of the results of literacy assessments conducted in South African primary schools. *South African Journal of Childhood Education*, *10*(1), 1-6. https://doi.org/10.4102/sajce.v10i1.745.

Graesser, A.C., Golding, J.M., & Long, D.L. (1991). Narrative representation and Comprehension. In R. Barr, M.L. Kamil, P. Mosenthal & P.D Pearson (Eds.), (pp.20-25), *Handbook of Reading Research*. Longman.

Gronmo, S. (2020). *Social Science Research: Qualitative, Quantitative and Mixed Methods Approaches*. SAGE publishers.

Hagell, P. (2014) Testing Rating scale Unidimensionality Using the Principal Component Analysis (PCA)/t-test Protocol with the Rasch Model: The Primacy

of Theory over Statistics, *Open Journal of Statistics, 4,* 456-459, http://dx.doi.org/10.4236/ojs.2014.46044.

Herron, J., & Gills, M. B. (2020). Encoding as a Route to Phoneme awareness and phonics: A shift in Literacy instruction. In M. B. Gills, & L. Maats, (Eds), The role of Phonology and Language in Learning to Read. (Vol.,46(3), pp. 17-25). International Dyslexia Association.

Hoang, V. V. (2006). *Translation: Theory and Practice- A textbook for Senior Student of English*. Vietnam Educational Publishing House.

Hosseini, A. (2021). The Differences Between Direct and Indirect Translation: Assessment of Two translations of the Japanese novel Black Rain. *Language Related Research, 11*(6), 1-4. https://lrr.modares.ac.ir/browse.php?a_id=38391&sid=14&slc_lang=en

Howie, S, J., Venter, E., Van Staden, S., Zimmerman, L., Long, C., Du Toit. C., Scherman, V., & Archer, E. (2006). *PIRLS 2006 Summary Report: South African children's reading literacy achievement*. Centre for Evaluation and Assessment.

Howie, S. J., Combrinck, C., Roux, K., Tshele, M., Mokoena, G. M., & Macleod-Palane, N. (2017). *PIRLS Literacy 2016: South African Children Reading Literacy Achievement.* Centre for Evaluation and Assessment.

Igwenagu, C. (2016). *Fundamentals of research methodology and data collection*. Lambert Academic Publishing.

Ing, M. L., Musah, M, B., Al-hudawi, S. H. V., Tahir, L. M., & Kamil, N. M. (2015). Validity of Teacher-Made Assessments: A table of Specification Approach. *Canadian Centre of Science and Education*, *11* (5), 194-195. http://dx.doi.org/10.5539/ass.v11n5p193.

Jettmarova, Z. (2012). Czech Translation Theory in Translation: A Review of Jiri Levy, The Art of Translation. *A Journal of Literary Studies and Linguistics, 2*(2). 183-187.

Johansone, I. (2023). Survey operations procedures in PIRLS 2021. In M. von Davier, I. V. S. Mullis, B. Fishbein, & P. Foy (Eds.), *Methods and Procedures: PIRLS 2021 Technical Report* (pp. 4.1-4.26). Boston College, TIMSS & PIRLS International Study Center.

Johansone, I., & Flicop, S. (2023). Quality assurance program for PIRLS 2021. In M. von Davier, I. V. S. Mullis, B. Fishbein, & P. Foy (Eds.), *Methods and Procedures: PIRLS 2021 Technical Report* (pp. 6.1-6.19). Boston College, TIMSS & PIRLS International Study Center.

Johnston, M. P. (2014). Secondary Data Analysis: A method of which the Time Has Come. *Qualitative and Quantitative Methods in Libraries*, *3*, 619-626. https://www.qqml-journal.net/index.php/qqml/article/view/169

Jones, E., & Kay, M. (1992). Instrumentation in cross-cultural research. *Nursing Research, 41*, 96–188. https://pubmed.ncbi.nlm.nih.gov/1584664/

Kadir, S. J., Zaim, M., & Refnaldi, R. (2019). Developing Instruments for Evaluating Validity, Practicality, and Effectiveness of the Authentic Assessment for Speaking Skill at Junior High School. *Advances in Social Science, Education and Humanities Research*, *276*, 98-99.https://doi.org/10.2991/icoelt-18.2019.14.

Karismawati, A. (2015). The Use of Indirect Translation Techniques in “The Hobbit: The Desolation of Smaug” Movie. [Bachelor of Education degree]. Semarang State University.

Kaya, S., & Ok. A. (2020). The Antecedents influencing the implementation and success of the middle school English language curriculum. *International Journal of Contemporary Educational Research*, *7*(1), 201-214. https://doi.org/10.33200/ijcer.660386.

Kembaren, F. W. R. (2018). *Translation Theory and Practice*. University of North Sumatra.

Killen, R. (2010). *Teaching Strategies Quality Teaching & Learning: Outcome-based assessment*. Juta Publishers.

Kim, M., Han, H., & Phillips, L. (2003). Metric Equivalence Assessments in Cross-Cultural Research: Using an Example of the Centre for Epidemiological Studies-Depression Scale. *Journal of Nursing Measurement*, *11*(1), 5-10. Doi:10.1891/106137403780954930.

Kirk, P. (2015). *The Importance of Literacy in a Child's development.* https://www.bizjournals.com/albanly/news/2015/01/21/the-importance-of-lit-in-a-childs.html

Kivinja, C., & Kuyini, B. A. (2017). Understanding and applying research paradigms in educational contexts. *Scieda Press*, *6*(5), 30-33. https://doi.org/10.5430/ijhe.v6n5p26

Korsnakova, P., Depts, S., Ebbs, D. (2020) Translation: The preparation of national language versions of assessment instrument. In H. Wagemaker (Eds.), Reliability and Validity of International Large-scale Assessments (Vol,10., pp 85-108) *Springer Publishers*.

Kubiszyn, T., & Borich, D. G. (2013). *Educational Testing and Measurement: Classroom Application and Practice*. Wiley and Sons Inc.

Labuschagne, M. J. (2015). *Validation of the performance of Tshivenda learners in PIRLS 2006.* [Masters Dissertation]. University of Pretoria.

Linacre, J. M. (1994). *Many-Facet Rasch Measurement*. (2nd ed.) Mesa Press.

Maree, K. (2019). *First Steps to Research.* (3rd ed.) Van Schaik Publishers

Markus, K., & Smith, K. (2010). Content Validity. In Salkind, N. (Eds.), *Encyclopaedia of research design* (pp. 108-109). SAGE Publications. https://dx.doi.org/10.4135/9781412961288.n74.

Martin, M. O., Mullis, I. V. S., & Hooper, M. (2017). *Methods and Procedures in PIRLS 2016*. Boston College, TIMSS & PIRLS International Study Center.

Martin, M. O., & Liu, J. (2019). PIRLS 2021 context questionnaire framework. In I. V. S. Mullis & M. O. Martin (Eds.), *PIRLS 2021 assessment frameworks* (pp. 27-56). Boston College, TIMSS & PIRLS International Study Center.

Mbiza, M. (2018). *The issues with South Africa's Education System*. https://fundiconnect.co.za/the-issues-with-south-africas-education-system/

McLeod Palane, N.L. (2017). *The Effect of Language of Instruction and Contextual Factors on Higher-order Reading Comprehension Performance.* [Doctoral Thesis]. University of Pretoria.

McMillian, J. H., & Schumacher, S. (2014). *Research in Education: Evidence-based Inquiry.* (7th ed.). Pearson Education.

Millsap, R. E. (2011). *Statistical Approaches to Measurement Invariance.* Routledge, Taylor and Francis group.

Mishra, B. S., & Alok, S. (2017). *Handbook of Research Methodology: A Compendium for Scholars and Researchers*. EDUcreation publishing.

Mohamed, S. (2020). *South Africa: Broken and Unequal Education Perpetuating Poverty and Inequality.* South African Publishers.

Motshekga, A. (2017). *2016 Progress in International Reading Literacy Study (PIRLS) Results.* Government publishers.

Motshekga, A. (2023). *National Seminar on reading literacy (PIRLS results*). Government publishers.

Mtsatse, N. (2017). *Exploring the Differential Item Functioning on reading achievement between English and IsiXhosa language subgroups.* [Masters Dissertation]. University of Pretoria.

Mullis, I. V. S., Martin, M. O., Kennedy, A., & Foy P. (2007). *PIRLS 2006 international report: IEA's Progress in International Reading Literacy Study in primary schools in 40 countries*. Boston College, TIMSS & PIRLS International Study Center

Mullis, I. V. S., & Martin, M. O. (2015). *PIRLS 2016 Assessment Framework.* Boston College, TIMSS & PIRLS International Study Center.

Mullis, I. V. S & Martin, M, O. (2021). *PIRLS 2021 Assessment Framework*. Boston College, TIMSS & PIRLS International Study Center.

Mullis, I. V. S., Martin, M. O., Foy, P & Hooper, M. (2017). *International Results in Reading*. Boston College, TIMSS & PIRLS International Study Center.

Mullis, I. V. S., Martin, M. O., Foy, P., & Hooper, M. (2017). *PIRLS 2021 Assessment Framework.* Boston College, TIMSS & PIRLS International Study Center.

National Centre for Education Statistics (NCES). (2020). Progress in International Reading Literacy Study (PIRLS). *National Centre for Education Statistics,* 1-10.

Nida, E. A. (1969). Science of Translation. *Linguistic Society of America, 45*(3), 483-498.

Nugraha, S.I. (2019). The Language in Education Policy in South Africa: A gap between policy and efficacy. *Advances in Social Sciences, Education and Humanities Research, 254*, 568-570. https://doi.org/10.2991/conaplin-18.2019.321.

Nugroho, B. A. (2007). Meaning and Translation. *Journal of English and Education*, *1* (2), 66-72. https://journal.uii.ac.id/JEE/article/view/6442

Null, W. (2011). *Curriculum: From theory to practise.* Rowman & Littlefield publishers.

Nwammuo, A. N., & Salawu, A. (2018). Media roles in disseminating strategies for teaching and learning indigenous languages: The case of South Africa's language-in-education policy in post-apartheid era, *Cogent Arts & Humanities, 5*(1), 1-13. https://doi.org/10.1080/23311983.2018.1553653.

Organisation for Economic Co-operation and Development (OECD). (2016). *PISA 2018 translation and adaptation guidelines.* National Project Managers' Meeting, Thailand, Prague, Czech.

Organisation for Economic Co-operation and Development. & Programme for International Student Assessment. (2005). *Contextual Framework for PISA 2006*.

Okeke, C., & Van Wyk, M. (2015). *Educational research: an African approach*. Oxford University Press.

Oluwatayo, J. (2012). Validity and reliability issues in educational research. *Journal of Educational and Social Research*, *2*(2), 393. https://doi.org/10.5901/jesr.2012.v2n2.391.

Peña, D. E. (2007). Lost in Translation: Methodological Considerations in Cross-Cultural Research. *Child Development*, *78*(4), 1255-1261. Doi:10.1111/j.1467-8624.2007. 01064.x.

Pham, L. (2018). A review of key paradigms: positivism, interpretivist, and critical inquiry. *Research gates publishers*, 2-3 https://doi.org.10.13140/RG.2.2./3995.54569.

Punch, K. (2011). *Introduction to social science research: Quantitative and qualitative approaches.* (4th ed.). SAGE.

Putnick, D. L., & Bornstein, M. H. (2016). Measurement invariance conventions and reporting: The state of the art and future directions for psychological research. *Developmental Review*, *41*, 71-79. http://dx.doi.org.10.1016/j.dr.2016.06.004.

Prah, K.K. (2006). *Challenges to the Promotion of Indigenous Languages in South Africa.* Review Commissioned by the Foundation for Human Rights in South Africa. *The Center for Advanced Studies of African Society.*

Rahman, S. (2016). The advantages and disadvantages of using Qualitative and Quantitative Approaches and methods in language: "Testing and assessment"

Research: A literature review. *Journal of education and learning*, *6*(1), 102-108 https://doi.org.10.5539/jel.v6nlp102.

Resnick, B. L., & Hampton, S. (2009). *Reading and Writing grade by grade* (Rev. ed.). New standards and International Reading Association.

Ronderos, N., Shavelson, R. J., Holtsch, D., Zlatkin-Troitschanskaia, O., & Solano-Flores, G. (2021). International Performance Assessment of Critical Thinking: Framework for Translation and Adaptation. *Journal Of Supranational Policies of Education, 13*, 62-87. https://doi.org/10.15366/jospoe2021.13.003.

Rosa, A. A., Pieta, H., & Maia, R. B. (2017). Theoretical, methodological and terminological issues regarding indirect translation: An overview. *Routledge Taylor & Francis Group*, 10(2), 113-114. http://dx.doi.org/10.1080/14781700.2017.1285247.

Roux, K. (2015). *Identifying the relationship between the home environment, parental attributes and learner achievement in reading literacy.* [Masters Dissertation]. University of Pretoria.

Roux, K. (2020). *Examining the Equivalence of the PIRLS 2016 Released Texts in South Africa Across Three Languages.* [Doctoral Thesis]. University of Pretoria.

Roux, K., Van Staden, S., Pretorius, E. J. (2022). Investigating the differential item function of a PIRLS Literacy 2016 text across three languages. *Journal of Education, 87*, 135-140. http://dx.doi.org/10.17159/2520-9868/i87a07.

Rule, P., & Land, S. (2017). Finding the plot in South African reading education. *Reading and Writing*, *8*(1), 1-8. https://doi.org/10.4102/rw.v8i1.121.

Sandilands, D., Oliveri, M. E., Zumbo, B. D., & Ercikan, K. (2013). Investigating sources of differential item functioning in international large-scale assessments using a confirmatory approach. *International Journal of Testing, 13*(2), 152–174. https://doi.org/10.1080/15305058.2012.690140.

Sefotho, M. M. (2018). *Philosophy in Education Research: African Perspectives*. Van Schaik Publishers.

Shah, R. S., & Al-Bargi, A. (2013). Research Paradigms: Researchers' Worldviews, Theoretical Framework and Study Design. *Arab World English Journal, International Peer Reviewed Journal*, *4*(4), 252-264. https://awej.org/images/AllIssues/Volume4/Volume4Number4Dec2013/15.pdf

Shea, M., & Ceprano, M. (2017). Reading with understanding: A global expectation. *Journal of inquiry & Action in Education,* *9*(1), 48-64. https://files.eric.ed.gov/fulltext/EJ1158259.pdf

Shiyab, S. M. (2013). Translation Quality Assessment: A perspective on pedagogy. *Arab World English Journal, International Peer Reviewed Journal, 2*, 40-45. https://awej.org/translation-quality-assessment-a-perspective-on-pedagogy/

Siddiek, G. A. (2010). The Impact of test content Validity on Language Teaching and Learning. *Asian Social Sciences: Canadian Centre of Science Education*, *6*(12), 1-10. https://files.eric.ed.gov/fulltext/ED574721.pdf

Singh, R. K.V., & Shaari, A. Z. (2019). The analysis of High-order Thinking skills in English reading comprehension tests in Malaysia. *Malaysian Journal of Society and Space, 15*(1), 12-19. https://doi.org/10.17576/geo-2019-1501-02.

Sireci, G. S. (2020). Standardization and Understandarzation in Educational Assessments *Educational Measurement: Issues and Practice*, *39*(3), 100-102. https://doi.org/10.1111/emip.12377.

Solano-Flores, G. (2012). Translation Accommodations Framework for Testing English Language Learners in Mathematics. *Smarter Balanced Assessment Consortium.* https://doi.org/portal.smarterbalanced.org/library/en/item-accessibility-and-languagevariation-conceptual-framework.pdf.

Solano-Flores, G., Backhoff, E., & Contreras-Niño, L. Á. (2009). Theory of Test Translation Error. *International Journal of Testing*, *9*(2), 78–91. https://doi.org/10.1080/15305050902880835.

Souza, M. A. P., Coster, W. J., Mancini, M. C., Dutra, F. C. M. S., Kramer, J., & Sampaaio, R. F. (2017). Rasch Analysis of the participation scale (P-scale): usefulness of the P-scale to a rehabilitation service network. *Bio Med Central Public health, 17*(934), 1-3. https://doi.org/10.1186/s12889-017-4945-9.

Spaull, N. (2013). Poverty & Privilege: Primary School Inequality in South Africa. *Journal of Educational Development 33*(5), 1-3. https://doi.org/10.1016/j.ijedudev.2012.09.009.

Spaull, N. (2021, September 13). *South Africa's "Right to Read and Write"*. Barksdale Reading Institute Resources. https://nicspaull.com/category/reading/#:~:text=The%20Right%20to%20Read%20and%20Write%3A,by%20the%20age%20of%2010.

Spicker, P. (2017). The Real Dependent Variable Problem: The Limitations of Quantitative Analysis in Comparative Policy Studies. *Social Policy & Administration*, *52*, 216-217. https://doi.org.10.1111/spol.12308.

Stemler, S. E., & Naples, A. (2021). Rasch Measurement vs. Item Response Theory: knowing when to cross the line. *Practical Assessment, Research and Evaluation*, *26*(11), 1-12. https://doi.org/10.7275/v2gd-4441.

Taherdoost, H. (2016). Validity and Reliability of Research Instruments; How to Test the Validation of a Questionnaire/Survey in a Research. *International Journal of Academic Research Management*, *5*(3), 29-33. http://dx.doi.org/10.2139/ssrn.3205040

Tanrikulu, L. (2017). *Translation Theories*. Veli University.

Tennant, A. (n.d,). *Reading Matter: What is reading?* MacMillian Education. https://www.onestopenglish.com/support-for-teaching-adults/reading-matters-what-is-reading/154842.article

The Republic of South Africa (RSA). (1996a). *The Bill of Rights of the Constitution of The Republic of South Africa. https://www.justice.gov.za/legislation/constitution/saconstitution-web-eng.pdf*.

The Republic of South African (RSA). (1996b). *The South African Schools Act 84 of 1996* (SASA). https://www.gov.za/sites/default/files/gcis_document/201409/act84of1996.pdf.

University of North Carolina (UNC). (2020). *Introduction to Visual Literacy.* University Libraries.

UNESCO. (n.d.). *The Global Alliance for Literacy within the framework of lifelong learning (GAL)*. UNESCO. https://unesdoc.unesco.org/ark:/48223/pf0000373033

Obdržálková, V. (2016) Translation as a decision-making process: an application of the model proposed by Jiri Levy to translation into a non-mother tongue. *Mutatis mutandis, 9*(2), 307-308.

Van de Schoot, R., Lugtig, P., Hox, J. (2012). A checklist for testing measurement invariance. *European Journal of Developmental Psychology, 9*(4), 486-490. http://doi.org/103389/fpsyg.2015.01064.

Van Staden, S. (2006). *PIRLS of wisdom: The What, Where, When and How of the International Reading Literacy Study in South Africa*. Centre for Evaluation and Assessment.

Van Staden, S. (2018). Exploring Possible Effects differential item Function of Reading Achievement Across Language Sub-groups: A South African Perspective. In L.D. Hill & F.J.F Levine. *Global Perspectives on Education Research.* (pp.9.1-9.22) Routledge.

Van Staden, S., & Roux, K. (2022). *South Africa*. In K.A. Reynolds, E. Wry, I.V.S. Mullis & M. von Davier, *PIRLS 2021 Encyclopedia*. https://pirls2021.org/south-africa.

Von Davier, M., Mullis, I.V.S., Fishbein, B., & Foy, P. (2023). *Methods and procedures: PIRLS 2021 Technical Report*. Boston College, TIMSS & PIRLS International Study Center.

Western Cape Government. (2020). *The importance of Reading.* Western Cape Government. https://www.westerncape.gov.za/general-publication/importance-reading

Wolhuter, C.C. (2017). The case of the education system of South Africa on the worl map: Inspiring case or call for guidance from abroad? In C.C. Wolhuter, L. Jacobs, & H.J. Steyn, (Eds.), Thinking about education systems (pp. 60-62), Keurkopie, Noordburg.

World Bank & United Nations Educational, Scientific and Cultural Organisations Institute for Statistics. (2021). *The Government Expenditure on Education, Total (% of Government Expenditure)*. https://data.worldbank.org/indicator/SE.XPD.TOTL.GB.ZS

Wry, E., & Mullis, I. V. S. (2023). Developing the PIRLS 2021 achievement instruments. In M. von Davier, I. V. S. Mullis, B. Fishbein, & P. Foy (Eds.), *Methods and Procedures: PIRLS 2021 Technical Report* (pp. 1.1-1.24). Boston College, TIMSS & PIRLS International Study Center.

Yin, L., Bezirhan, U., Fishbein, B., & Foy, P. (2023). Implementing the PIRLS 2021 achievement scaling methodology. In M. von Davier, I. V. S. Mullis, B. Fishbein, & P. Foy (Eds.), *Methods and Procedures: PIRLS 2021 Technical Report* (pp. 11.1-11.82). Boston College, TIMSS & PIRLS International Study Center.

Zanon, C., Hutz, S, C., Yoo, H. H., & Hambleton, K. R. (2016). An Application of Item Response Theory to psychological test development. *Psicologia: Reflexao e Critica*, *29*(18), 1-10. https://10.1186?s41155-016-0040-x.

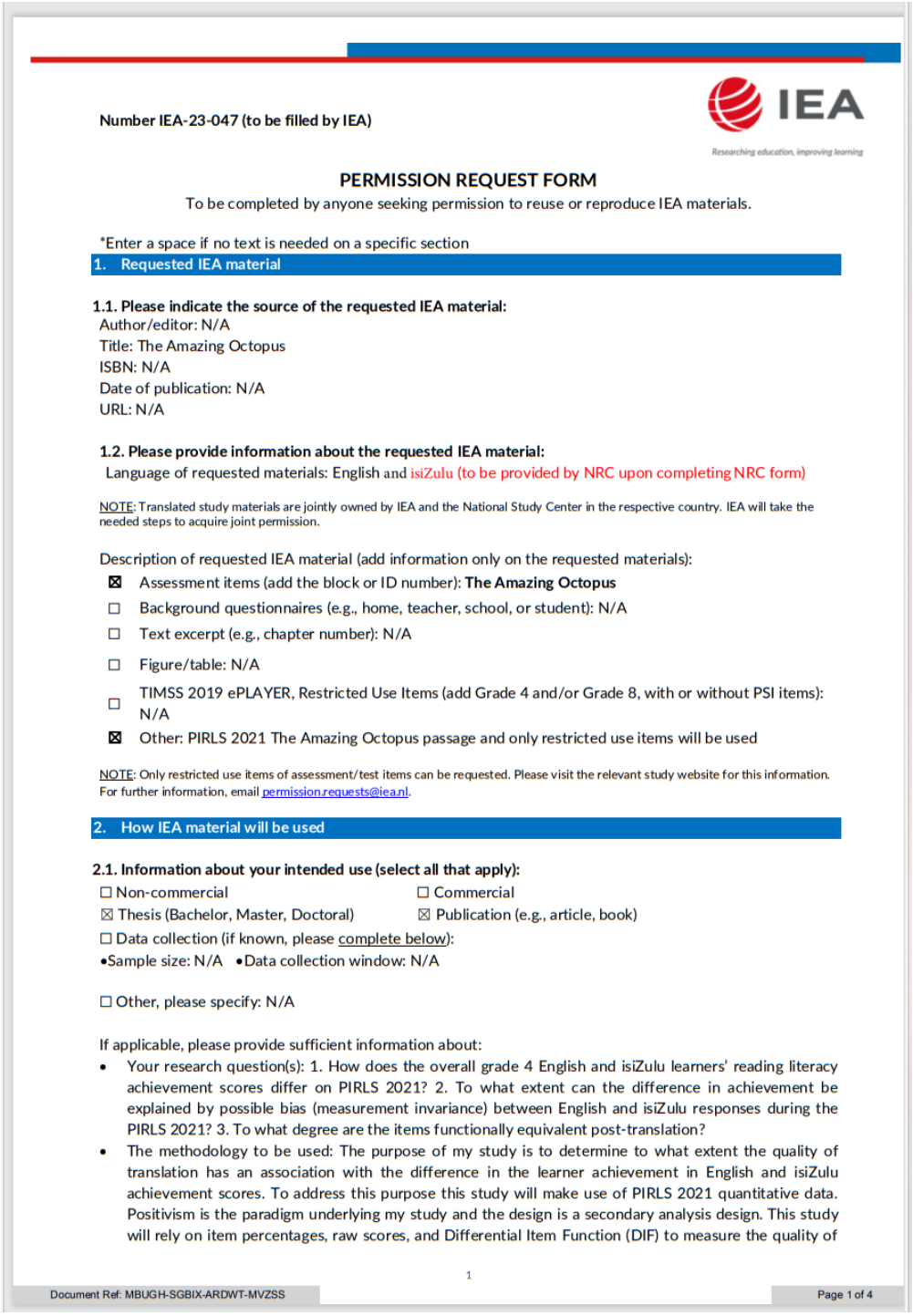
Zhao, X., & Zhu, L. (2012). Schema Theory and College English Reading Teaching. *Canadian Centre of Science and Education*, *5*(11), 111-112. Doi:10.5539/elt.v5n11p111.

Zimmerman, L. (2010). *The influence of schooling conditions and teaching practices on curriculum implementation for Grade 4 learner reading literacy development.* [Doctoral Thesis]. University of Pretoria.

Zumbo, B. (1999). A handbook on the theory and methods of differential item functioning: Logistic regression modelling as a unitary framework for binary and Likert-type item scores. *Directorate of Human Resource Research and Evaluation*.

APPENDIXES

Appendix A: International Associations for the Evaluation of Educational Achievement Permission Request Form



Number IEA-23-047 (to be filled by IEA)

IEA
Researching education, improving learning

PERMISSION REQUEST FORM

To be completed by anyone seeking permission to reuse or reproduce IEA materials.

*Enter a space if no text is needed on a specific section

1. Requested IEA material

1.1. Please indicate the source of the requested IEA material:
Author/editor: N/A
Title: The Amazing Octopus
ISBN: N/A
Date of publication: N/A
URL: N/A

1.2. Please provide information about the requested IEA material:
Language of requested materials: English and isiZulu (to be provided by NRC upon completing NRC form)

NOTE: Translated study materials are jointly owned by IEA and the National Study Center in the respective country. IEA will take the needed steps to acquire joint permission.

Description of requested IEA material (add information only on the requested materials):

- ☒ Assessment items (add the block or ID number): **The Amazing Octopus**
- ☐ Background questionnaires (e.g., home, teacher, school, or student): N/A
- ☐ Text excerpt (e.g., chapter number): N/A
- ☐ Figure/table: N/A
- ☐ TIMSS 2019 ePLAYER, Restricted Use Items (add Grade 4 and/or Grade 8, with or without PSI items): N/A
- ☒ Other: PIRLS 2021 The Amazing Octopus passage and only restricted use items will be used

NOTE: Only restricted use items of assessment/test items can be requested. Please visit the relevant study website for this information. For further information, email permission.requests@iea.nl.

2. How IEA material will be used

2.1. Information about your intended use (select all that apply):
☐ Non-commercial ☐ Commercial
☒ Thesis (Bachelor, Master, Doctoral) ☒ Publication (e.g., article, book)
☐ Data collection (if known, please complete below):
•Sample size: N/A •Data collection window: N/A

☐ Other, please specify: N/A

If applicable, please provide sufficient information about:

- Your research question(s): 1. How does the overall grade 4 English and isiZulu learners' reading literacy achievement scores differ on PIRLS 2021? 2. To what extent can the difference in achievement be explained by possible bias (measurement invariance) between English and isiZulu responses during the PIRLS 2021? 3. To what degree are the items functionally equivalent post-translation?
- The methodology to be used: The purpose of my study is to determine to what extent the quality of translation has an association with the difference in the learner achievement in English and isiZulu achievement scores. To address this purpose this study will make use of PIRLS 2021 quantitative data. Positivism is the paradigm underlying my study and the design is a secondary analysis design. This study will rely on item percentages, raw scores, and Differential Item Function (DIF) to measure the quality of

1

Document Ref: MBUGH-SGBIX-ARDWT-MVZSS Page 1 of 4

Number IEA-23-047 (to be filled by IEA)

IEA
Researching education, improving learning

PERMISSION REQUEST FORM

To be completed by anyone seeking permission to reuse or reproduce IEA materials.

*Enter a space if no text is needed on a specific section

1. Requested IEA material

1.1. Please indicate the source of the requested IEA material:

Author/editor: N/A
Title: The Amazing Octopus
ISBN: N/A
Date of publication: N/A
URL: N/A

1.2. Please provide information about the requested IEA material:

Language of requested materials: English and isiZulu (to be provided by NRC upon completing NRC form)

NOTE: Translated study materials are jointly owned by IEA and the National Study Center in the respective country. IEA will take the needed steps to acquire joint permission.

Description of requested IEA material (add information only on the requested materials):

- ☒ Assessment items (add the block or ID number): **The Amazing Octopus**
- ☐ Background questionnaires (e.g., home, teacher, school, or student): N/A
- ☐ Text excerpt (e.g., chapter number): N/A
- ☐ Figure/table: N/A
- ☐ TIMSS 2019 ePLAYER, Restricted Use Items (add Grade 4 and/or Grade 8, with or without PSI items): N/A
- ☒ Other: PIRLS 2021 The Amazing Octopus passage and only restricted use items will be used

NOTE: Only restricted use items of assessment/test items can be requested. Please visit the relevant study website for this information. For further information, email permission.requests@iea.nl.

2. How IEA material will be used

2.1. Information about your intended use (select all that apply):

☐ Non-commercial ☐ Commercial
☒ Thesis (Bachelor, Master, Doctoral) ☒ Publication (e.g., article, book)
☐ Data collection (if known, please complete below):
•Sample size: N/A •Data collection window: N/A

☐ Other, please specify: N/A

If applicable, please provide sufficient information about:

- Your research question(s): 1. How does the overall grade 4 English and isiZulu learners' reading literacy achievement scores differ on PIRLS 2021? 2. To what extent can the difference in achievement be explained by possible bias (measurement invariance) between English and isiZulu responses during the PIRLS 2021? 3. To what degree are the items functionally equivalent post-translation?
- The methodology to be used: The purpose of my study is to determine to what extent the quality of translation has an association with the difference in the learner achievement in English and isiZulu achievement scores. To address this purpose this study will make use of PIRLS 2021 quantitative data. Positivism is the paradigm underlying my study and the design is a secondary analysis design. This study will rely on item percentages, raw scores, and Differential Item Function (DIF) to measure the quality of

Number IEA-23-047 (to be filled by IEA)

IEA
Researching education, improving learning

translation and thereafter determine whether there is a measurement invariance between the English and isiZulu achievement scores. My study will use SPSS in conjunction with IDB-Analyser for descriptive statistics and use RUMM2030 for DIF analysis.

- The intended use of data and plans for reporting it: The results will be used in Masters Dissertation chapters, journal article publications and conference presentations.

2.2. Information about where the requested materials or research results will appear:

- ☒ Print publication
- ☐ Website (URL:N/A)
- ☒ Online publication (e.g., online journal)
- ☒ Other, please specify: Print and online version of the dissertation

2.3. Additional information about the planned publication (if applicable):

Author: Sinethemba Mthimkhulu
Title: Reading comprehension of isiZulu grade 4 home language learners: Translation validity in PIRLS 2021
Publisher or sponsor: University of Pretoria
Language of publication: English
Intended audience: Scholars and educators
Number of copies for distribution: Less than 10
Retail price: N/A
Release date: N/A
Additional comments: The dissertation will be submitted to the library and be available online on the university's website.

3. Requestor information

First name: Sinethemba
Last name: Mthimkhulu
Name of institution or organization: University of Pretoria
City and zip code: Pretoria, 0001
Country: South Africa
Email: u21727393@tuks.co.za Mthimkhulu SB

Signature:  - Mthimkhulu S Date of request:25 April 2023

To submit your permission request, please sign and return this completed form to IEA via email (permission.requests@iea.nl). By signing, you agree that the requested material will be used only as part of, or alongside the original work, where the primary value does not lie with the requested material itself.
Please note that by signing this form you confirm that you have filled out this form truthfully and to the best of your knowledge. You also confirm that you have read and will comply with the conditions. Providing IEA with incorrect or incomplete information will not only invalidate permission if granted but can also hold you liable for any damage arising from your failure to comply with these requirements. In case you have any hesitations and/or reservations regarding the information you have to fill out on this form, please contact IEA via email to permission.requests@iea.nl. **Please allow 4-5 weeks** for the processing of your permission request. Please note that your personal information (name, institution, and email address) provided on this form will be archived by IEA for documentation purposes.

Disclaimer

Please note that the website and its contents, together with all online and/or printed publications and restricted use items ('works') by TIMSS, PIRLS, ICCS, ICILS, and other IEA studies, were created with the utmost care. However, the correctness of the information cannot be always guaranteed. IEA cannot and will not be held responsible or liable for any damages that may arise from the use of these resources, nor will IEA be liable for the wrongful use and/or interpretation of its works.

Number IEA-23-047 (to be filled by IEA)

IEA
Researching education, improving learning

Please be advised that IEA cannot authorize the use of texts or items that include third-party copyrighted materials (e.g., reading passages in PIRLS, photographs, images). Users of any third-party copyrighted materials must first seek and be granted copyright permission from the owner of the content as indicated in the copyright citation line.

Please note that permission is only granted for the case as described in this form. Any additional use of this or any other IEA materials requires further permission. IEA copyright must be explicitly acknowledged, and the need to obtain permission for any further use of the published text/material must be clearly stated in the requested use/display of this material. IEA, its proprietary assessment instruments, and studies are all the result of the choices and combination of elements by which the creator has expressed its creativity in an original matter, further to which a result has been achieved which is an intellectual creation and therefore protected as a copyright protected work, as stipulated in article 10 of the Dutch Copyright Act ("DCA") and article 2(a) of Directive 2001/29/EC regarding the harmonization of certain aspects of copyright and related rights in the information society (the "EU Copyright Directive").[1] The copyrights in these works are owned by IEA. National versions of instruments are recognized as the joint venture and shared intellectual property of IEA and the relevant participating institutions and should be treated accordingly.

IEA has a strict Intellectual Property Policy in place regarding third-party use of its copyright protected instruments and studies. All publications and restricted use items of TIMSS, PIRLS, and other IEA studies, as well as translations thereof, are for non-commercial, educational, and research purposes only. Prior permission is required when using IEA data sources for assessments or learning materials. As stated, IEA reserves the right to refuse copy deemed inappropriate or not properly sourced. IEA Intellectual Property Policy is *inter alia* included on the IEA website (http://www.iea.nl) and on the TIMSS and PIRLS website (https://timssandpirls.bc.edu/), in which it is clearly stated that all accessible instruments and/or data are IEA proprietary copyright protected. Said webpages also contain links to its permission form, which should be submitted with IEA prior to any use of its materials and/or instruments.

TIMSS, PIRLS, ICCS, and ICILS are registered trademarks of IEA. Use of these trademarks without permission of IEA by others may constitute trademark infringement. Furthermore, the website and its contents, together with all online and/or printed publications and restricted use items by TIMSS, PIRLS, ICCS, ICILS, and other IEA studies are and will remain copyright of IEA.

Exploitation, distribution, redistribution, reproduction, and/or transmitting in any form or by any means, including electronic or mechanical methods such as photocopying, information storage, and retrieval system of these publications, restricted use items, translations thereof and/or part thereof are prohibited unless written permission has been provided by IEA.

When quoting and/or citing from one or more publications and/or restricted use items from TIMSS, PIRLS, ICCS, ICILS, and other IEA studies for the sake of educational or research purposes, please print an acknowledgment of the source, including the year and name of that publication and/or restricted use item. Please use the following acknowledgment as an example:

SOURCE: TIMSS 2007 Assessment. Copyright © 2009 International Association for the Evaluation of Educational Achievement (IEA). Publisher: TIMSS & PIRLS International Study Center, Lynch School of Education, Boston College.

Please note that citing without naming the source can or will constitute plagiarism for which you can and will be held accountable.

To be completed by IEA

4. Permission

☒ **Granted** ☐ **Denied**

Terms of agreement: Permission is granted for non-exclusive rights to reproduce the material requested above upon the terms and solely for the purpose indicated.

☒ Free ☐ Agreed Fee (payable to NL63ABNA0481961968)

Signature: _____ Date: 2023-07-24

Name: Dirk Hastedt

[1] Cf. ECJ 16 July 2009, Case C-5/08 (Infopaq I).

3

Document Ref: MBUGH-SGBIX-ARDWT-MVZSS

Page 3 of 4

Number IEA- 23-047 (to be filled by IEA)

Researching education, improving learning

Title: Executive Director

Signature Certificate

Reference number: MBUGH-SGBIX-ARDWT-MVZSS

Signer	Timestamp	Signature
Dirk Hastedt Email: d.hastedt@iea.nl		
Sent:	24 Jul 2023 12:44:13 UTC	
Viewed:	24 Jul 2023 14:18:41 UTC	
Signed:	24 Jul 2023 14:19:31 UTC	
Recipient Verification:		
✓Email verified	24 Jul 2023 14:18:41 UTC	IP address: 212.123.229.166 Location: Amsterdam, Netherlands

Document completed by all parties on:
24 Jul 2023 14:19:31 UTC

Page 1 of 1

Signed with PandaDoc

PandaDoc is a document workflow and certified eSignature solution trusted by 40,000+ companies worldwide.

Appendix B: Department of Basic Education Permission Request Form

PERMISSION REQUEST FORM

To be completed by anyone seeking permission to reuse or reproduce the Department of Basic Education (DBE) material.

*Enter a space if no text is needed on a specific section.

1. Requested DBE material

1.1. Please indicate the source of the requested DBE material:
Author/editor: N/A
Title: The Amazing Octopus passage and items
ISBN: N/A
Date of publication: N/A
URL: N/A

1.2. Please provide information about the requested DBE material:
Language of requested materials: English and isiZulu

Description of requested DBE material (add information only on the requested materials):

- ☑ Assessment items (add the block or ID number): The Amazing Octopus passage and items
- ☐ Background questionnaires (e.g., home, teacher, school, or student): N/A
- ☐ Text excerpt (e.g., chapter number): N/A
- ☐ Figure/table: N/A
- ☐ Other:

NOTE: Only restricted use items of assessment/test items can be requested. Please visit the relevant study website for this information.For further information, email chetty.m@dbe.gov.za.

2. How DBE material will be used

2.1. Information about your intended use (select all that apply):

- ☐ Non-commercial
- ☐ Commercial
- ☑ Thesis (Bachelor, Master, Doctoral)
- ☑ Publication (e.g., article, book)
- ☐ Data collection (if known, please complete below): N/A
  - Sample size:
  - Data collection window: N/A
- ☐ Other, please specify: N/A

If applicable, please provide sufficient information about:

- Your research question(s):

How does the overall grade 4 English and isiZulu learners' reading literacy achievement scores differ

1

on PIRLS 2021? 2. To what extent can the difference in achievement be explained by possible bias (measurement invariance) between English and isiZulu responses during the PIRLS 2021? 3. To what degree are the items functionally equivalent post-translation?

- The methodology to be used:

The purpose of my study is to determine to what extent the quality of translation has an association with the difference in the learner achievement in English and isiZulu achievement scores. To address this purpose this study will make use of PIRLS 2021 quantitative data. Positivism is the paradigm underlying my study and the design is a secondary analysis design. This study will rely on item percentages, raw scores, and Differential Item Function (DIF) to measure the quality of translation and thereafter determine whether there is a measurement invariance between the English and isiZulu achievement scores. My study will use SPSS in conjunction with IDB-Analyser for descriptive statistics and use RUMM2030 for DIF analysis.

- The intended use of data and plans for reporting it:

The results will be used in Masters Dissertation chapters, journal article publications and conference presentations.

2.2. Information about where the requested materials or research results will appear:

☑ Print publication
☑ Online publication (e.g., online journal)
☐ Website
☑ Other, please specify: Conference proceedings

2.3. Additional information about the planned publication (if applicable):

Author: Mthimkhulu Sinethemba
Title: Examining PIRLS 2021: Differential item function across English and isiZulu
Language of publication: English
Intended audience: Researchers, Teachers, Student-teachers
Release date: 2023
Additional comments: N/A

3. Requestor information

First name: Sinethemba
Last name: Mthimkhulu
Name of institution or organization: University of Pretoria
City and Zip Code: Pretoria, 0028
Country: South Africa
Email: u21727393
Signature: ***Mthimkhulu SB***

Date of request: 21/08/2023

To submit your permission request, please sign and return this completed form to DBE via email (chetty.m@dbe.gov.za). By signing, you agree that the requested material will be used only as part of, or alongside the original work, where the primary value does not lie with the requested material itself.
Please note that by signing this form you confirm that you have filled out this form truthfully and to the best of your knowledge. You also confirm that you have read and will comply with the conditions. Providing DBE with incorrect or incomplete information will not only invalidate permission if granted but can also hold you liable for any damage arising from your failure to comply with these requirements. In case you have any hesitations and/or reservations regarding the information you have to fill out on this form, please contact DBE via email to chetty.m@dbe.gov.za. Please allow 4-5 weeks for the processing of your permission request. Please note that your personal information (name, institution, and email address) provided on this form will be archived by DBE for documentation purposes.

Disclaimer

Please note that the website and its contents, together with all online and/or printed publications and restricted use items ('works') by TIMSS, PIRLS, SE, ELNA, and other DBE studies, were created with the utmost care. However, the correctness of the information cannot be always guaranteed. DBE cannot and will not be held responsible or liable for any damages that may arise from the use of these resources, nor will DBE be liable for the wrongful use and/or interpretation of its works.

Please be advised that DBE cannot authorise the use of texts or items that include third-party copyrighted materials (e.g., reading passages in PIRLS, photographs, images). Users of any third-party copyrighted materials must first seek and be granted copyright permission from the owner of the content as indicated in the copyright citation line.

Please note that permission is only granted for the case as described in this form. Any additional use of this or any other DBE materials requires further permission. DBE copyright must be explicitly acknowledged, and the need to obtain permission for any further use of the published text/material must be clearly stated in the requested use/display of this material. DBE, its proprietary assessment instruments, and studies are all the result of the choices and combination of elements by which the creator has expressed its creativity in an original matter, further to which a result has been achieved which is an intellectual creation and therefore protected as a DBE copyright protected work. The copyrights in these works are owned by DBE. National versions of instruments are recognised as the joint venture and shared intellectual property of DBE and the relevant participating institutions and should be treated accordingly.

DBE has a strict Intellectual Property Policy in place regarding third-party use of its copyright protected instruments and studies. All publications and restricted use items of TIMSS, PIRLS, and other DBE studies, as well as translations thereof, are for non-commercial, educational, and research purposes only. Prior permission is required when using DBE data sources for assessments or learning materials. As stated, DBE reserves the right to refuse copy deemed inappropriate or not properly sourced. DBE Intellectual Property Policy is inter alia included on the DBE website (www.dbe.gov.za) in which it is clearly stated that all accessible instruments and/or data are DBE proprietary copyright protected. The DBE website and its contents, together with all online and/or printed publications and restricted use items by TIMSS, PIRLS, SE, ELNA, and other DBE studies are and will remain copyright of DBE.

Exploitation, distribution, redistribution, reproduction, and/or transmitting in any form or by any means, including electronic or mechanical methods such as photocopying, information storage, and retrieval system of these publications, restricted use items, translations thereof and/or part thereof are prohibited unless written permission has been provided by DBE.

When quoting and/or citing from one or more publications and/or restricted use items from TIMSS, PIRLS, SE,

ELNA, and other DBE studies for the sake of educational or research purposes, please print an acknowledgment of the source, including the year and name of that publication and/or restricted use item.

Please note that citing without naming the source can or will constitute plagiarism for which you can and will be held accountable.

To be completed by DBE

4.Permission

☑ Granted ☐ Denied

Terms of agreement: Permission is granted for non-exclusive rights to reproduce the material requested above upon the terms and solely for the purpose indicated.

☑ Free
☐ Agreed Fee (Payable to the Department of Basic Education)

Name: Dr Mark Chetty

Signature: 

Date: 21/8/2023

Position: NATIONAL RESEARCH COORDINATOR - TIMSS & PIRLS

4